(54) Title: POLYPEPTIDES

(57) Abstract: A polypeptide (interacting polypeptide) capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises the amino acid sequence PP(T/N)K and is less than 150 amino acids in length or is not full-length Xenopus or human FAST1 or a fragment thereof, mouse FAST2, Xenopus Milk, Xenopus Mixer or Xenopus Bix2. The Smad polypeptide may be Smad2 or Smad3. The interacting polypeptides are useful in screening assays and in medicine.

IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

1
POLYPEPTIDES

The present invention relates to polypeptides and polynucleotides and their use in medicine and screening methods.

5

Members of the TGF-β superfamily are secreted signalling molecules produced by cells to influence the behaviour of their neighbours, by regulating cell proliferation, survival, adhesion, differentiation and specification of developmental fate (Hogan *et al* 1994; Kingsley, 1994; Massagué, 1998). These

10 ligands bind a type II receptor which allows transphosphorylation of the type I receptor (Massagué, 1998). This in turn leads to phosphorylation and activation of the receptor-activated class of Smads (R-Smads; Massagué, 1998), which are responsible for transducing signals from the activated receptors to the nucleus. Smad proteins are a family of highly conserved, intracellular proteins that signal

15 cellular responses downstream of transforming growth factor-beta (TGF-beta) family serine/threonine kinase receptors. R-Smads 2 and 3 are phosphorylated by TGF-β or activin type I receptors, whilst R-Smads 1, 5 and 8 are substrates for BMP type I receptors (Massagué, 1998). Phosphorylation relieves an auto-inhibitory interaction of the C-terminal MH2 domain with the N-terminal MH1

20 domain (Hata *et al* 1997), allowing the R-Smads to form heteromeric complexes *via* their MH2 domains with members of the Smad4 class (Lagna *et al* 1996; Zhang *et al* 1997; Masuyama *et al* 1999; Howell *et al* 1999). These activated complexes translocate to the nucleus to regulate transcription of target genes (Whitman, 1998). Smads bind DNA very weakly alone (Shi *et al* 1998) and are

25 primarily recruited to DNA by other DNA-binding transcription factors (Derynck *et al* 1998; Whitman, 1998), the prototype being the winged-helix/forkhead transcription factor, Fast-1 (Chen *et al* 1996; Chen *et al* 1997).

2

These co-operating transcription factors are likely to be key determinants of cell type specificity of TGF-β signaling, but are mostly still poorly characterized.

5   The *Xenopus* embryo provides an excellent system in which to elucidate the basis of specificity in TGF-β signaling pathways. In the *Xenopus* embryo, TGF-β family members act as morphogens, playing key roles in the patterning of different tissues (Green and Smith, 1990; Gurdon *et al* 1994; Hogan, 1996; Whitman, 1998). For example, an activin-like signal, which requires the maternal transcription factor VegT for its production, is released by the vegetal hemisphere of the embryos to

10   induce mesoderm in the overlying equatorial cells (Harland and Gerhart, 1997; Kimelman and Griffin, 1998; Zhang *et al* 1998). The same signaling molecule is also thought to be responsible for specifying endoderm (Henry *et al* 1996). Patterning of the mesoderm and endoderm depends on the precise transcriptional responses of cells within the prospective meso-endoderm to this signal. But what

15   determines which genes are induced in response to this activin-like signal in particular cells, and how is their expression maintained? The presence of particular transcription factors that cooperate with Smads in some cells, but not others could obviously play an important role, as could the presence of other cooperating signaling pathways such as Wnt, FGF and BMP (reviewed by Harland and

20   Gerhart, 1997; Heasman, 1997; Whitman, 1998). The existence in *Xenopus* embryos of multiple transcription factors which are capable of recruiting activin-activated Smads and have different DNA-binding specificity has been proposed, based on the fact that the activin-responsive elements defined in the promoters of differentially expressed meso-endodermal genes share little sequence similarity

25   (reviewed in Howell and Hill, 1997).

3

The mechanism that confines expression of the Xenopus *goosecoid* gene (Blumberg *et al* 1991) to the dorsal marginal zone of the early gastrula embryo is beginning to be understood. It results from a synergistic interaction between a Wnt signal acting through a proximal element (PE) in the promoter (Watabe *et al* 1995;

5   Laurent *et al* 1997) and an activin-like signal acting through a distal element (DE). The DE is also conserved in the mouse and zebrafish *goosecoid* promoters (Watabe *et al* 1995; Candia *et al* 1997; McKendry *et al* 1998). Since the sequence of the DE bears no resemblance to the ARE from the *Mix.2* promoter, the transcription factors involved in its activin-inducibility may be distinct from Fast-

10   1. A paired-like homeodomain factor of unknown identity has been implicated in the activin-responsive transcription of the DE-related element in the zebrafish *goosecoid* promoter (McKendry *et al* 1998).

The TGFβ superfamily, signalling pathways and likely functions have been

15   extensively researched and reviewed. TGFβ appears to be involved in the modulation of many biological processes and may be implicated in pathogenic conditions including tumour growth, inflammation, wound healing, scarring, fibrosis, kidney damage, for example in diabetes, and atherosclerosis. Proteins related to TGFβ include activins, inhibins and bone morphogenetic proteins

20   (BMPs). In some situations, enhancement of TGFβ signalling may be beneficial, whilst in others, inhibition may be useful. A lack of specific small-molecule agonists or antagonists of TGFβ signalling has impeded investigations, particularly *in vivo*.

25   TGFβ appears to play two contradictory roles in tumorigenesis (reviewed in Akhurst & Balmain (1999) "Genetic events and the role of TGFβ in epithelial tumour progression." *J Pathol* 187, 82-90). At early stages of tumorigenesis it

4

acts as a tumour supressor through its ability to growth arrest epithelial cells, from which approximately 90% of human tumours are derived. However, at late stages of tumorigenesis, TGFβ is a powerful tumour promoter, acting directly on the tumour cells themselves, promoting malignant conversion and

5   tumour invasion, and acting indirectly by promoting angiogenesis and immunosuppression. Inhibition of the ability of TGFβ to act as a tumour promoter without affecting its antiproliferative responses may therefore be desirable.

10   The views expressed in a selection of reviews are summarised below.

Hartsough MT; Mulder KM (1997) "Transforming growth factor-β signalling in epithelial cells" *Pharmacol Ther* **75** (1), 21-41 discusses the resistance of some tumours to growth suppression by TGFβ.

15

Noble NA; Border WA (1997) "Angiotensin II in renal fibrosis: should TGF-β rather than blood pressure be the therapeutic target?" *Semin Nephrol* 17(5), 455-66 discusses the role of TGFβ in promoting tissue fibrosis and the induction of TGFβ by angiotensin II.

20

Koli K; Keski-Oja J (1996) "Transforming growth factor-β system and its regulation by members of the steroid-thyroid hormone superfamily." *Adv Cancer Res* **70**, 63-94, discusses TGF-βs and their receptors and their action as key regulators of many aspects of cell growth, differentiation, and function,

25   particularly malignancy.

5

Grande JP (1997) "Role of transforming growth factor-β in tissue injury and repair." *Proc Soc Exp Biol Med* 214(1), 27-40 discusses the role of TGFβ in normal cell growth, development, and tissue remodelling following injury. Disruption of the TGFβ1 gene *in utero* produces a wasting syndrome

5   characterised by systemic inflammation, suggesting that this growth factor plays an important role in limiting the inflammatory response. TGFβ is a dominant mediator of the pathologic extracellular matrix accumulation that characterises progression of tissue injury to end-stage organ failure. Recent studies directed towards characterisation of the TGFβ genes, dissection of the mechanisms by

10   which TGFβs are produced and activated, and identification of TGFβ signalling pathways have established the important roles that these family members play in cell and tissue homeostasis. TGFβ structure-function relationships and their relevance to models of tissue injury/wound repair are also discussed.

15   Lawrence DA (1996) "Transforming growth factor-β: a general review." *Eur Cytokine Netw* 7(3), 363-74 reviews the roles of TGF-β1, β2 and β3 in mammals. The author comments that they play critical roles in growth regulation and development. All three of these growth factors are secreted by most cell types, generally in a latent form, requiring activation before they can

20   exert biological activity. This activation of latent TGF-β, which may involve plasmin, thrombospondin and possibly acidic microenvironments, appears to be a crucial regulatory step in controlling their effects. The TGF-βs possess three major activities: they inhibit proliferation of most cells, but can stimulate the growth of some mesenchymal cells; they exert immunosuppressive effects; and

25   they enhance the formation of extracellular matrix. Two types of membrane receptors (type I and type II) possessing a serine/threonine kinase activity within their cytoplasmic domains are involved in signal transduction. Inhibition of

6

growth by the TGF-βs stems from a blockage of the cell cycle in late G1 phase. Among the molecular participants concerned in G1-arrest are the Retinoblastoma (Rb) protein and members of the Cyclin/Cyclin-dependent kinase/Cyclin dependent kinase inhibitor families. In the intact organism the TGF-βs are

5   involved in wound repair processes and in starting inflammatory reactions and then in their resolution. The latter effects of the TGF-βs derive in part from their chemotactic attraction of inflammatory cells and of fibroblasts. From gene knockout and from overexpression studies it has been shown that precise regulation of each isoform is essential for survival, at least in the long term.

10   Several clinical applications for certain isoforms have already shown their efficacy and they have been implicated in numerous other pathological situations.

Pignatelli M; Gilligan CJ (1996) "Transforming growth factor-β in GI neoplasia, wound healing and immune response." *Baillieres Clin Gastroenterol* **10(1)**, 65-

15   81 discusses the influence that cell-cell and cell-matrix interactions, the differentiating status of the cell together with the functional activity of other soluble growth factors have on responses to TGF-βs, particularly in relation to homeostasis of the GI mucosa and their role in gastrointestinal carcinogenesis.

20   Cox DA (1995) "Transforming growth factor-β 3." *Cell Biol Int* **19(5)**, 357-71 discusses the molecular and cellular biology of TGF-β 3 and those physiological actions which may lead to clinical applications, particularly in the indication areas of wound healing and chemoprotection.

25   Wahl SM (19920 "Transforming growth factor β (TGF-β) in inflammation: a cause and a cure." *J Clin Immunol* **12(2)**, 61-74 discusses the mechanisms controlling whether the pro- or antiinflammatory effects of this peptide prevail.

7

Ruscetti FW; Palladino MA (1991) "Transforming growth factor-β and the immune system." *Prog Growth Factor Res* 3(2), 159-75 discusses the increased levels of TGF-β found in several disease states associated with immunosuppression such as different forms of malignancy, chronic degenerative diseases, and AIDS, implicating the involvement of TGF-β in the pathogenesis of some diseases.

TGFβ is known to be an <u>inhibitor</u> of inflammation (as reviewed, for example, in Lawrence (1996) and Grande (1997), both cited above) for example from studies in which massive inflammatory lesions are seen in mice in which a TGFβ gene is inactivated.

Here we identify new partners for activated Smads. We have identified a short motif, characterized by containing the preferred sequence PP(T/N)K, that appears to be necessary and may be sufficient for interaction with the MH2 domain of Smad2. Full-length Smad polypeptides, for example Smad2 and Smad3, may be activated by phosphorylation near the C-terminus of the polypeptide, which induces a conformational change which exposes a binding site in the MH2 domain for transcription factors such as FAST1, FAST2 or the newly-identified partners. A Smad polypeptide in which the N-terminal domain is not present or is truncated may not require phosphorylation in order to expose this binding site.

A first aspect of the invention provides a polypeptide (interacting polypeptide) capable of interacting with a Smad polypeptide wherein the interacting

8

polypeptide comprises a Smad Interaction Motif (SIM) and is less than 32, 31, or 30 amino acids in length.

The interacting polypeptide/SIM preferably comprises the amino acid sequence

5   PP(T/N)K or three out of four residues thereof. The three residues may be any three residues; ie the three residues need not be consecutive residues. Thus, the interacting polypeptide/SIM may comprise the amino acid sequence PPSK or PPQK ie a residue with an aliphatic hydroxyl side chain or an amide side chain may be present between the PP and K residues. It is strongly preferred that an

10   alanine residue is not present instead of the T/N residue, ie between the PP and K residues.

It is not essential for all residues of the putative PP(N/T)K motif to be correct, as noted above and discussed further in Example 2 and the legend to Figure 15.

15   For example, either of the proline residues may be replaced, for example by an alanine residue, or the lysine residue may be replaced, for example by an alanine. The two prolines may not be of equal importance in that mutation of the second P to A appears to have a larger impact on the ability to bind Smad2C and activate transcription.   An order of preference for these residues may

20   therefore be PP>A(for example)P> >PA(for example). At least one P may be important in this position ie pair of residues; in the case of mixer, the PP-AA mutant does not appear to bind to Smad2C.

As discussed further below, the residue immediately before (ie N-terminal of)

25   the amino acids corresponding to the sequence motif PP(T/N)K may preferably be a hydrophobic residue, for example F, M or V. The residue immediately after (ie C-terminal of) the amino acid sequence corresponding to the sequence

9

motif PP(T/N)K ie at position +1 may preferably be an S or T, which may be immediately followed by an I or V residue. An acidic residue (for example glutamate or aspartate) may be present at position about +3 to about +10, preferably +4 or +5 and may be immediately followed by a hydrophobic residue, for example M, V or I. A proline residue may be present at a position from 5 to about 20 residues C-terminal of the amino acid sequence corresponding to the sequence motif PP(T/N)K.

An acidic residue (for example glutamate or aspartate) immediately followed by a hydrophobic residue (for example F, Y, L) may be present at position starting about –20 or -17 to –2 relative to the amino acids corresponding to the PP(T/N)K sequence motif, preferably at –9 to –8 or –5 to –4 or –2 to –1 (ie immediately N-terminal of the PP(T/N)K sequence motif. A leucine residue may be present at position about –2 to –15, preferably about –5 to –10. The leucine residue may be the hydrophobic residue that is immediately preceded by an acidic residue, as noted above.

Thus, the SIM (and polypeptide) may comprise at least 8, 9 or 10 (preferably 10 or 11) of the specified residues (ie not residues designated by an X) of the amino acid sequence D/E-Hyd-$(X)_n$-P-P-(N/T)-K-(T/S)-(I/V)-$(X)_m$-(D/E)-(M/V/I)-$(X)_k$-P

wherein m = 0 to 7; k = 0 to 8 or 12; n = 0 to 15 or 18.

It will be appreciated that this motif may extend over a stretch of more than 32, 31 or 30 amino acids. It is preferred that a SIM, for example conforming to this motif, extends over a stretch of 32, 31, 30, 29, 28, 27, 26, 25 or fewer amino acids but it will be appreciated that this is not essential.

10

A polypeptide of less than 32 amino acids in length may comprise a SIM and be capable of interacting with a Smad polypeptide without comprising all elements of the D/E-Hyd-$(X)_n$-P-P-(N/T)-K-(T/S)-(I/V)-$(X)_m$-(D/E)-(M/V/I)-$(X)_k$-P motif;

5 for example the polypeptide may be a polypeptide as defined below and in claim 12 which does not have residues corresponding to the D/E-Hyd-$(X)_n$ residues, because the N-terminal amino acids of the polypeptide correspond to the PPNK motif (for example a polypeptide consisting of the amino acid sequence PPNKTITPDMNVRIPPI).

10

It is preferred that there is a leucine residue at position about –2 to –15, preferably about –5 to –10 N-terminal of the residues corresponding with the PP(N/T)K sequence; the leucine residue may be the hydrophobic residue that is immediately preceded by an acidic residue (ie D/E).

15

It is preferred that a residue which does not match the consensus sequence indicated above is an alanine residue.

By "interacting with" is included the meaning of "binding to", for example

20 detectably binding to, for example binding detectable using any method of detecting protein/protein binding as indicated below, for example co-immunoprecipitation or a surface plasmon resonance technique. The term "polypeptide" in connection with the interacting polypeptide includes peptides as small as the peptide PPNK or PPTK. The invention includes a polypeptide of

25 less than 32, 31 or 30 amino acids in length comprising the amino acid sequence PP(T/N)K.

11

A further aspect of the invention provides a polypeptide (interacting polypeptide) capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises a SIM, for example the amino acid sequence PP(T/N)K, or three out of four residues thereof, and is not full-length *Xenopus* or human

5   FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer, *Xenopus* Bix3, Bix2 or Bix 1. As discussed below, Bix1 is not an interacting polypeptide. It may be preferred that the interacting polypeptide is not Zebrafish FAST1 or Zebrafish Mixer. The interacting polypeptide may be *Xenopus* FAST3, the sequence of which is shown in Figure 13.

10

The terms FAST1, FAST2, Milk, Mixer, Bix3, Bix2 and Bix1 are well known to those skilled in the art. Mixer may also be known as Mix3 (see Mead *et al* (1998)). The sequence for FAST2 is given, for example, in Liu *et al* (1999) *Mol Cell Biol* **19**, 424-430 or Labbé *et al* (1998) *Mol Cell* **2**, 109-120. The sequence

15   for FAST1 is given, for example, in Chen *et al* (1996) *Nature* **383**, 691-696 and Chen *et al* (1997) *Nature* **389**, 85-89. The sequence of human Fast1 is given in Zhou *et al* (1998) *Mol Cell* **2**, 121-127 and in WO98/5380. Fragments of FAST1 are described in Chen *et al* (1997) *Nature* **389**, 85-89 and in WO98/5380. The sequence for Milk is given in Ecochard *et al* (1998). The

20   sequence for Mixer is given in Henry & Melton (1998). The sequences for Bix3, Bix2 and Bix1 are given in Tada *et al* (1998). Bix1 may also be known as Mix4 (see Mead *et al* (1998) Cloning of Mix-related homeodomain proteins using fast retrieval of gel shift activities, (FROGS), a technique for the isolation of DNA-binding proteins *Proc Natl Acad Sci U S A* **95(19)**, 11251-6).

25   Accession numbers for these polypeptides are listed below:

Fast Family Members   Accession number

12

Xenopus Fast-1	U70980 (Chen, X et al (1996) Nature 383, 691-696)
Xenopus Fast3	See Figure 13 and Figure 18
Zebrafish Fast-1	AF263000
Human Fast-1	AF076292 (Zhou et al (1998) Mol. Cell 2, 121-127)
5 Mouse Fast-2	AF069303 (Labbé et al (1998) Mol. Cell 2, 109-120)

Mix family members	Accession number
Xenopus Mix.1	M27063 (Rosa (1989) Cell 57: 965-974.)
10	
Xenopus Mix.2	U50745 (Vize (1996) Dev. Biol. 177, 226-231)
Xenopus Mixer	AF068263 (Henry and Melton (1998) Science 281, 91-96)
15	
Xenopus Milk	AF005999 (Ecochard et al (1998) Development 125, 2577-2585)
Xenopus Bix1	AF079559 (Tada et al (1998) Development 125, 3997-4006)
20	
Xenopus Bix3	AF079561 (Tada et al (1998))
Xenopus Bix4	AF079562 (Tada et al (1998))
Zebrafish Mixer	AF121771 (Alexander et al (1999) Dev. Biol. 215, 343-357)
25	
Chick Mix	U34615 (Peale et al (1998) Mech. Dev. 75, 179-182)

13

Mouse Mix   AF135063 (Pearce & Evans (1999) *Mech. Dev.* 87,

189-192)

5   Note that the sequence of the gene called Bix 2 (accession number AF079560 is

virtually identical to Milk and it is most probably the same gene as Milk.

The interacting polypeptide may be a transcription factor or a fragment thereof.

Thus, the interacting polypeptide may comprise a domain that is capable of

10   binding to a nucleic acid, preferably DNA, still more preferably double-stranded

DNA, yet more preferably to DNA that forms part of a promoter region for a

gene. The interacting polypeptide may be a fragment of a transcription factor

wherein the transcription factor comprises a said domain that is capable of

binding to a nucleic acid but the interacting polypeptide does not comprise the

15   said domain. It will be appreciated that the interacting polypeptide may bind to

the said nucleic acid with higher affinity when the interacting polypeptide is

bound to one or more other polypeptides, for example one or more Smad

polypeptides, than when it is not so bound. The interacting polypeptide may

bind to the said nucleic acid as a dimer or as a heterodimer with another

20   transcription factor ie with another polypeptide comprising a domain that is

capable of binding to a nucleic acid. The interacting polypeptide may be capable

of promoting transcription of DNA; additional polypeptides may be required for

transcription to take place. The interacting polypeptide may comprise, for

example, a winged-helix DNA binding domain or a Paired DNA binding domain

25   or a homeodomain, for example a Paired-like homeodomain. It will be

appreciated that the interacting polypeptide may comprise more than one domain

that is capable of binding to a nucleic acid.

14

As is well known to those skilled in the art, a promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. A promoter may be a region of DNA capable of

5   controlling transcription of neighbouring DNA. It will be appreciated that a transcription factor that is capable of interacting with a Smad polypeptide may not be capable of binding to DNA unless it is in a complex with the Smad polypeptide, for example Smad2 or Smad3 and Smad4. The transcription factor may be capable of interacting (directly or indirectly) with an RNA polymerase.

10   It is preferred that the transcription factor is capable of interacting directly with an RNA polymerase.

FAST1 and FAST2 comprise a winged-helix (also known as a Forkhead) DNA binding domain. Members of the Mix family, which may include the chicken

15   CMIX polypeptide (Peale *et al* (1998) *Mech of Dev* **75**, 167-170 and Stein *et al* (1998) *Mech of Dev* **75**, 163-165), comprise a Paired-like homeodomain (see, for example Wilson *et al* (1993)*Genes Dev* **7**, 2120-2134).

The term paired homeodomain transcription factor is well known to those skilled

20   in the art. Paired homeodomain transcription factors are reviewed in, for example, Galliot *et al* (1999) Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class *Dev Genes Evol* **209**, 186-197, Wright *et al* (1989) Vertebrate homeodomain proteins: families of region-specific transcription factors *Trends Biochem Sci* **14**, 52-56 and Dorn *et al* (1994)

25   Homeodomain proteins in development and therapy *Pharmacol Ther* **61**, 155-184.

15

The homeobox domain has about 60 amino acids and consists of a helix-turn-helix motif that binds DNA by inserting the recognition helix into the major groove of the DNA and its amino-terminal arm into the adjacent minor groove. Representative homeobox domain are found in the *Drosophila* Antennapedia polypeptide and the *Drosophila* Paired polypeptide.

Galliot *et al* (1999) *Dev Genes Evol* 209, 186-197 reviews polypeptides belonging to the Paired class. This class of polypeptides contain a homeobox DNA binding domain that is related to that found in the *Drosophila* gene Paired *(prd)* and characterised by invariant residues which distinguish them from other homeodomain (HD) classes. Three subclasses can be defined according to the residue at position 50 of the homeodomain, which plays a key role in determining DNA binding specificity. The Pax or Prd-type genes have a serine residue at position 50 (S_{50} type) and also have a second DNA-binding domain, the prd (Paired) domain. Mammalian members of this sub-class include the Pax genes (see, for example, Adams *et al* (1992) *Genes & Dev* 6, 1589-1607). A second sub-class has a lysine at position 50 (K_{50} type) and a third sub-class has a glutamine residue (Q_{50} type) at position 50. The K_{50} and Q_{50} sub-classes do not have the prd domain. The Mix family of polypeptides belongs to the Q_{50} class.

The paired domain motif is a domain of 128 amino acids identified as a secondary homology region in the homeobox-containing proteins of the *Drosophila paired* and *gooseberry* genes (Bopp *et al* (1986) *Cell* 47, 1033-1040; Baumgartner *et al* (1987) *Genes & Dev* 1, 1247-1267). The paired domain motif encodes a DNA-binding motif (Goulding *et al* (1991) *EMBO J* 10, 1135-1147; Treisman *et al* (1991) *Genes & Dev* 5, 594-604; Chalepakis *et al* (1991) *Cell* 66, 873-884). Three α-helices are predicted to be present in the paired domain (see

16

Bopp *et al* (1989) *EMBO J* 8, 3447-3457). The paired domain proteins of vertebrates are encoded by a multigene family that has been conserved in evolution, termed the Pax gene family, as mentioned above.

5   The term Forkhead or winged helix polypeptide is well known to those skilled in the art. Forkhead/winged helix polypeptides are reviewed, for example, in Kaufmann & Knochel (1996) *Mech Dev* 57, 3-20. A polypeptide may be identified as a Forkhead or winged-helix polypeptide if it comprises a domain with features of a Forkhead/winged-helix DNA binding domain. The

10   Forkhead/winged-helix domain is a variant of the helix-turn-helix motif (Brennan (1993) The winged-helix DNA-binding motif: Another helix-turn-helix takeoff *Cell* 74, 773-776; Clark *et al* (1993) Co-crystal structure of the HNF-3/forkhead DNA-recognition motif resembles histone H5 *Nature* 364, 412-420). The forkhead/winged-helix domain is responsible for DNA-binding specificity

15   and binds to DNA as a monomer, with two loops or wings on the C-terminal side of the helix-turn-helix.

The forkhead domain is about 111 amino acids in length. Based on the degree of homology within the forkhead domain, the forkhead family is further split into

20   subgroups. Over 80 genes with the conserved wing-helix forkhead motif have been identified from yeast to mammalian sources, as reviewed in Kaufmann & Knochel (1996) *Mech Dev* 57, 3-20. Sequence identity in the 111 amino acid domain may be more than about 50%, for example between about 70% and 95% identity; sequence identity outside this domain between forkhead family members

25   may be much less. A Forkhead protein may have at least 30, 40, 50, 60, 75, 80, 85, 90 or 95% amino acid sequence identity with the FKHR Forkhead domain (Davis *et al* (1995) *Hum Mol Genet* 4, 2355-2362).

17

The Forkhead domains of FAST1 and FAST2 are about 40% identical to that of HNF-3β and several other family members (Liu *et al* (1999). FAST1 and FAST2 are highly homologous in the Forkhead domain and have sequence

5   similarity in other domains. No homology to other Forkhead proteins is observed outside the Forkhead domain. FAST polypeptides therefore appear to form a sub-family of the Forkhead family.

The sequence of a novel FAST polypeptide, termed *Xenopus* FAST 3 is shown

10   in Figure 13. The nucleotide sequence is shown in Figure 18.

It will be appreciated that the interacting polypeptide may bind Smad2 and/or Smad3 MH2 domains but may not bind Smad1 or Smad4 directly. The interaction may require the α-helix 2 of the MH2 domain, though the interaction

15   may not be with the α-helix 2. The interaction may require regions equivalent to the regions of Smad2 indicated in Table 1 to be required for the interactions investigated.

It is preferred that the Smad polypeptide with which the interacting polypeptide

20   interacts is Smad2 or Smad3, more preferably human Smad2 or human Smad3, most preferably human Smad2. The terms Smad, Smad2 and Smad3 are well know to those skilled in the art; see, for example Massagué (1998); Macias-Silva *et al* (1996) *Cell* 87, 1215-1224 (human Smad2); Graff *et al* (1996) *Cell* 85, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* 383, 168-172 (human Smad3).

25   The sequence of *Xenopus* Smad3, a novel Smad polypeptide, is shown in Figure 12 with the sequences of human Smads 2 and 3 and *Xenopus* Smad2.

18

Accession numbers for further Smad2 and Smad3 polypeptides are indicated below. Note that the α-helix2 region of these Smad2 and Smad3s, which is the region required for interaction with the SIM, is absolutely conserved with the previously characterized Xenopus Smad2 and human Smad2.

Drosophila Smad2   AF101386 (Brummel *et al* (1999) *Genes Dev.* 13, 98-111)

Zebrafish Smad2   AF229022 (Dick *et al* (2000) *Gene* 246, 69-80)

Chick Smad2 fragment   AF230190

Chick Smad3 fragment   AF230191

It will be appreciated that a Smad polypeptide may have a domain recognisable as an MH2 domain. The MH2 domains of *Drosophila*, *Xenopus*, human and mouse Smad2, for example, appear to be more than 90% identical (Brummel *et al* (1999) *Genes Dev* 13, 98-111). A tryptophan residue may be present at the residue equivalent to W274 of *Xenopus* Smad2 (see, for example, WO97/22697). Smads 1, 2, 3, 4, 5 and 8 may further have a conserved domain recognisable as a MH1 domain, whilst Smads 6 and 7 may have a divergent MH1 domain. Smads 2 and 3 may be activated by TGFβ or activin by phosphoryation at two serine residues near the C-terminus of the polypeptide. Smads 2 and 3 may be cytoplasmic until activated and then translocate to the nucleus. Smads 2 and 3 may also form a complex with Smad4 in response to ligand.

In terms of sequence, Smad2 and 3 may be defined by the sequence in the L3 loop which may dictate their binding to the activin and TGFβ type I receptors and the sequence of the α-helix 2 that is required to bind to Fast1 (see Shi *et al* (1997) *Nature* 388, 87-93 and WO99/01765), Milk and Mixer (see below).

19

The Smad polypeptide may be a variant, fragment, derivative or fusion of human Smad2 or human Smad3.

5   It is preferred that the Smad polypeptide has a greater amino acid identity with the C-terminal MH2 region, particularly the α-helix2 region (see Chen *et al* (1998)), of Smad2 or Smad3, for example human Smad2 or Smad3, than with the C-terminal MH2 region, particularly the α-helix2 region, of Smad1 or Smad4, for example human Smad1 or human Smad4. The MH2 domain of

10   *Xenopus* Smad2 starts at amino acid W-274.

By "variants" of a polypeptide, for example of Smad2 or Smad3, we include insertions, deletions and substitutions, either conservative or non-conservative. In particular we include variants of the polypeptide where such changes do not

15   substantially alter the activity of the said polypeptide, for example the ability of the Smad polypeptide to bind to an interacting polypeptide, for example a transcription factor such as FAST1, FAST2, Mixer or Milk, or another Smad polypeptide, for example Smad4.

20   By "conservative substitutions" is intended combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

It is particularly preferred if the Smad polypeptide variant has an amino acid sequence which has at least 65% identity with the amino acid sequence of Smad2

25   or Smad3, for example the amino acid sequence of Smad2 or Smad3 shown in Macias-Silva *et al* (1996) *Cell* 87, 1215-1224 (human Smad2); Graff *et al* (1996) *Cell* 85, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* 383, 168-172

(human Smad3) or Figure 12 (*Xenopus* Smad3), more preferably at least 50%, 55%, 60%, 70%, still more preferably at least 75%, yet still more preferably at least 80%, in further preference at least 85%, in still further preference at least 90% and most preferably at least 95% or 97% identity with the amino acid

5   sequence defined above.

It is still further preferred if the Smad polypeptide variant has an amino acid sequence which has at least 65% identity with the amino acid sequence of the α-helix2 domain of Smad2 or Smad3 shown in Macias-Silva *et al* (1996) *Cell* 87,

10   1215-1224 (human Smad2); Graff *et al* (1996) *Cell* 85, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* 383, 168-172 (human Smad3) or Figure 12 (*Xenopus* Smad3), more preferably at least 70% or 73%, still more preferably at least 75%, yet still more preferably at least 80%, in further preference at least 83% or 85%, in still further preference at least 90% and most preferably at least

15   95% or 97% identity with the amino acid sequence defined above. It will be appreciated that the α-helix2 domain of a Smad polypeptide may be readily identified by a person skilled in the art and as described in Chen *et al* (1998), for example using sequence comparisons as described below.

20   The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally.

25

21

The alignment may alternatively be carried out using the Clustal W program (Thompson *et al* (1994) *Nucl Acid Res* 22, 4673-4680). The parameters used may be as follows:

Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent.

Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05.

Scoring matrix: BLOSUM.

"Variations" of the polypeptide also include a polypeptide in which relatively short stretches (for example 5 to 20 amino acids) have a high degree of homology (at least 80% and preferably at least 90 or 95%) with equivalent stretches of the polypeptide even though the overall homology between the two polypeptides may be much less. This is because important active or binding sites may be shared even when the general architecture of the protein is different.

It is preferred that the Smad polypeptide, for example Smad2 or Smad3 polypeptide is a polypeptide which consists of the amino acid sequence of the Smad2 or Smad3 polypeptide as shown in Macias-Silva *et al* (1996) *Cell* 87, 1215-1224 (human Smad2); Graff *et al* (1996) *Cell* 85, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* 383, 168-172 (human Smad3) or Figure 12 (*Xenopus* Smad3), or naturally occurring allelic variants thereof and fusions thereof. A preferred fusion may be a GST fusion, for example as described in Example 1 or any other fusion described in Example 1 or a Myc fusion as described, for example, in Chen *et al* (1997). A further preferred fusion may have the tag Glu-Phe-Met-Pro-Met-Glu (termed EE-tag) or a His, HA or FLAG tag, as well known to those skilled in the art.

22

Alternatively, it is preferred that the Smad polypeptide is a fragment or a fusion of a fragment of a Smad2 or Smad3 polypeptide, as shown in Macias-Silva *et al* (1996) (human Smad2); Graff *et al* (1996) (*Xenopus* Smad2); Zhang *et al* (1996)

5   (human Smad3) or Figure 12 (*Xenopus* Smad3), or naturally occurring allelic variants thereof. It is preferred that the said fragment or fusion of a fragment comprises the MH2 domain, in particular the α-helix 2 domain of the said Smad2 or Smad3 polypeptide, as shown in the references indicated above, or naturally occurring allelic variants thereof. Particularly preferred fragments or

10   fusions include the fragments indicated in Table 1 as capable of binding to the endogenous activity, Mixer, Milk or Fast-1, and fusions of those fragments, for example with GST.

It is preferred that the Smad polypeptide is a polypeptide that is capable of

15   binding to FAST1, FAST2, FAST3, Mixer, Milk, or Bix3. The capability of the said Smad polypeptide with regard to binding FAST1, FAST2, FAST3, Mixer, Milk, or Bix3 may be measured by any method of detecting/measuring a protein/protein interaction, as discussed further below and in Example 1. Suitable methods include yeast two-hybrid interactions, co-purification (for

20   example co-immunoprecipitation or GST-pulldown assays), ELISA, co-immunoprecipitation methods and bandshift assays.

It will be appreciated that it may be necessary for the Smad polypeptide to be phosphorylated in order for FAST1, FAST2, Mixer, Milk, or Bix3 or the said

25   interacting polypeptide, for example FAST3, to be capable of binding to the Smad polypeptide ie for the Smad polypeptide to be activated. Phosphorylation of a full-length Smad polypeptide may be necessary to relieve an auto-inhibitory

23

interaction of the C-terminal MH2 domain with the N-terminal MH1 domain, as discussed above. Smad fragments in which the N-terminal MH1 domain is absent, disrupted or truncated may not require phosphorylation in order for the interacting polypeptide to interact with the fragment. The relevant

5   phosphorylation of Smad2 takes place on residues Ser465 and Ser467 (see, for example, Souchelnytskyi *et al* (1997) *J Biol Chem* **272**, 28107-28115). Phosphorylation may be performed *in vitro*, for example by immunoprecipitating active recepetor complexes from Cos1 cells overexpressing the receptors and treated with TGFβ. These immunoprecipitates will phosphorylate GST-Smad2,

10   for example as described in Macias-Silva *et al* (1996) *Cell* **87**, 1215-1224. It is preferred that the Smad polypeptide is a polypeptide, for example a fragment in which the N-terminal MH1 domain is absent, disrupted or truncated, that does not require phosphorylation in order to be able to bind to FAST1, FAST2, Mixer, Milk, Bix 2 or 3 or the said interacting polypeptide, for example FAST3.

15   Suitable Smad polypeptides may be the preferred fragments and fusions capable of binding to the endogenous activity, Mixer, Milk or FAST1 listed in Table 1.

The interacting polypeptide may be capable of interacting with a portion of the Smad polypeptide that is equivalent to α-helix 2 or part thereof of a full length

20   Smad polypeptide, for example Smad2 or Smad3.

It is preferred that the interacting polypeptide or PP(T/N)K- containing polypeptide is less (in order of preference) than 150, 100, 80, 70, 55, 50, 45, 40, 35, 32, 31, 30, 28 or 26 amino acids in length. It is further preferred that

25   the interacting polypeptide is at least (in order to preference) 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 28 or 30 amino acids in length or any combination of these maximum and minimum lengths. It is particularly

24

preferred that the interacting polypeptide is between 4 and about 30, 33 or 35 amino acids in length; in further preference the interacting polypeptide is between 25 and about 30, 33 or 35 amino acids in length.

5   It is preferred if the interacting polypeptide consists of a fragment of a naturally occurring protein such as those described below or a fusion thereof. Suitably, the fragment of a naturally occurring protein is less than (in order of preference) 150, 100, 80, 70, 55, 50, 45, 40, 35, 33, 32, 31, 30, 28 or 26 amino acids in length. Also suitably, the fragment of a naturally occurring protein is at least (in

10   order of preference) 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 28, 30 or 33 amino acids in length.

As indicated above, it is preferred that the interacting polypeptide further has an acidic (ie negatively charged) amino acid residue present at a position from 3 to

15   10, preferably 4 to 5 residues C-terminal of the amino acid sequence corresponding to the PP(T/N)K motif (and may be immediately followed by a hydrophobic residue, for example M, V or I), and/or a proline residue present at a position from 5 to 20 residues C-terminal of the amino acid sequence corresponding to the PP(T/N)K motif, as discussed above. The acidic

20   (negatively charged) amino acid residue is typically a glutamate or aspartate residue. At least one of the two proline residues within the PP(T/N)K motif are believed to be essential for interaction with the Smad polypeptide, as discussed above and further in relation to Figure 15 below. Polypeptides with the sequences AANK or QTNK in place of PP(T/N)K appear not to bind Smad2.

25   The downstream proline and acid (for example aspartate) residues as described above may also be important for binding. The residue immediately before (ie N-terminal of) the amino acid sequence PP(T/N)K may preferably be a

25

hydrophobic residue, for example F, M or V. The residue immediately after (ie C-terminal of) the amino acid sequence PP(T/N)K ie at position +1 may preferably be an S or T, which may be immediately followed by an I or V residue.

5

An acidic residue (for example glutamate or aspartate) immediately followed by a hydrophobic residue (for example F, Y, L) may be present at position starting about –20 to –2 relative to the amino acids corresponding to the PP(T/N)K sequence motif, preferably at –9 to –8 or –5 to –4 or –2 to –1 (ie immediately N-

10   terminal of the PP(T/N)K sequence motif. A leucine residue may be present at position about –2 to –15, preferably about –5 to –10. The leucine residue may be the hydrophobic residue that is immediately preceded by an acidic residue, as noted above.

15   It is particularly preferred that the interacting polypeptide consists of or comprises the amino acid sequence PPNKTITPDMNVRIPPI or PPNKTITPDMNTIIPQI or PPNKSVFDVLTSHPGD or PPNKSIYDVWVSHPRD or PPNKSIYDVWVSHPRD or PPNKTVFDIPVYTGHPG or PPNKTITPDMNTIIPQI or

20   PPNKTIGPEMKVVIPPL or PPNKSSKRGNTPPW or LLMDFNNFPPNKTITPDMNVRIPPI or HSNLMMDFPPNKTITPDMNTIIPQI or LDNMLRAMPPNKSVFDVLTSHPGD or LDSLFQGVPPNKSIYDVWVSHPRD or LDALFQGVPPNKSIYDVWVSHPRD or

25   LKNAPSDFPPNKTVFDIPVYTGHPG or HSNLVMEFPPNKTITPDMNTIIPQI or LVEYDNFPPNKTIGPEMKVVIPPL or ITSDAYSDSCPPPNKSSKRGNTPPW.

26

The interacting polypeptide may consist of or comprise the amino acid sequence of residues 283 to 307 of *Xenopus* Mixer, residues 316 to 340 of *Xenopus* Milk, residues 470 to 493 of *Xenopus* FAST1, residues 363 to 386 of mouse FAST2, residues 316 to 341 of *Xenopus* Bix2, resiudes 305 to 319 of *Xenopus* Bix 3, residues 327 to 350 of human FAST1, residues 363 to 386 of human FAST1, residues 245 to 269 of *Xenopus* FAST3, or the equivalent residues of the equivalent mammalian, preferably human, Mixer, Milk, Bix2/3, FAST1, FAST2 or FAST3 polypeptides or zebrafish polypeptides, for example zebrafish FAST1 or Mixer.

The interacting polypeptide or PP(T/N)K-containing polypeptide typically comprises the amino acid sequence X_n[SIM; for example PP(T/N)K] Z_m wherein X_n represents the amino acid sequence of the consecutive n amino acids immediately N terminal to the SIM (for example amino acid sequence PP(T/N)K) in a naturally occurring polypeptide comprising a SIM, for example the amino acid sequence PP(T/N)K, for example a said naturally occurring polypeptide described above, and wherein Z_m represents the amino acid sequence of the consecutive m amino acids immediately C terminal to the SIM, for example immediately C terminal to the amino acid sequence PP(T/N)K, in a naturally occurring polypeptide comprising the SIM, for example comprising the amino acid sequence PP(T/N)K, for example a said naturally occurring polypeptide described above, wherein n and m may independently be any number between 0 and 1, 5, 10, 15, 20, 25, 30, 50, 80, 100, 150, 200, 300 or 500 amino acids, preferably between 0 and 150, still more preferably between 0 and 30 amino acids. It is preferred that the amino acid sequences X_n and Z_m are immediately N and C terminal, respectively, to the SIM, for example the amino acid sequence PP(T/N)K, in the same naturally occurring polypeptide.

27

By "residue equivalent to" a particular residue, for example the residue Pro291 of full-length *Xenopus* Mixer, is included the meaning that the amino acid residue occupies a position in the native two or three dimensional structure of a polypeptide, for example a transcription factor comprising a Paired-like homeodomain, corresponding to the position occupied by the said particular residue, for example Pro291, in the native two or three dimensional structure of full-length *Xenopus* Mixer. It will be appreciated that Pro291 of *Xenopus* full-length Mixer is located outside the Paired-like homeodomain, towards the C-terminus of the polypeptide.

The residue equivalent to a particular residue, for example the residue Pro291 of full-length *Xenopus* Mixer, may be identified by alignment of the sequence of the polypeptide with that of full-length *Xenopus* Mixer in such a way as to maximise the match between the sequences. The alignment may be carried out by visual inspection and/or by the use of suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group, which will also allow the percent identity of the polypeptides to be calculated. The Align program (Pearson (1994) in: Methods in Molecular Biology, Computer Analysis of Sequence Data, Part II (Griffin, AM and Griffin, HG eds) pp 365-389, Humana Press, Clifton). Thus, residues identified in this manner are also "equivalent residues".

It will be appreciated that in the case of truncated forms of Mixer or in forms where simple replacements of amino acids have occurred it is facile to identify the "equivalent residue".

28

The sequence for *Xenopus* Mixer is given in, for example, Henry & Melton (1998).

The three-letter and one-letter amino acid code of the IUPAC-IUB Biochemical Nomenclature Commission is used herein. The sequence of polypeptides are given N-terminal to C-terminal as is conventional. In particular, Xaa represents any amino acid. It is preferred that the amino acids are L-amino acids, in particular it is strongly preferred that the SIM, for example a PP(T/N)K motif, consists of L-amino acid residues. It is preferred that the amino acid residues immediately flanking (such as those within 10 to 20 residues) of the SIM, for example flanking the PP(T/N)K motif are L-amino acids residues, but they may be D-amino acid residues.

The above polypeptides or peptide may be made by methods well known in the art and as described below and in Example 1, for example using molecular biology methods or automated chemical peptide synthesis methods.

Peptides may be synthesised by the Fmoc-polyamide mode of solid-phase peptide synthesis as disclosed by Lu *et al* (1981) *J. Org. Chem.* **46**, 3433 and references therein. Temporary N-amino group protection is afforded by the 9-fluorenylmethyloxycarbonyl (Fmoc) group. Repetitive cleavage of this highly base-labile protecting group is effected using 20% piperidine in N,N-dimethylformamide. Side-chain functionalities may be protected as their butyl ethers (in the case of serine threonine and tyrosine), butyl esters (in the case of glutamic acid and aspartic acid), butyloxycarbonyl derivative (in the case of lysine and histidine), trityl derivative (in the case of cysteine) and 4-methoxy-2,3,6-trimethylbenzenesulphonyl derivative (in the case of arginine). Where glutamine

29

or asparagine are C-terminal residues, use is made of the 4,4'-dimethoxybenzhydryl group for protection of the side chain amido functionalities. The solid-phase support is based on a polydimethyl-acrylamide polymer constituted from the three monomers dimethylacrylamide (backbone-monomer), bisacryloylethylene diamine (cross linker) and acryloylsarcosine methyl ester (functionalising agent). The peptide-to-resin cleavable linked agent used is the acid-labile 4-hydroxymethyl-phenoxyacetic acid derivative. All amino acid derivatives are added as their preformed symmetrical anhydride derivatives with the exception of asparagine and glutamine, which are added using a reversed N,N-dicyclohexyl-carbodiimide/1-hydroxybenzotriazole mediated coupling procedure. All coupling and deprotection reactions are monitored using ninhydrin, trinitrobenzene sulphonic acid or isotin test procedures. Upon completion of synthesis, peptides are cleaved from the resin support with concomitant removal of side-chain protecting groups by treatment with 95% trifluoroacetic acid containing a 50% scavenger mix. Scavengers commonly used are ethanedithiol, phenol, anisole and water, the exact choice depending on the constituent amino acids of the peptide being synthesised. Trifluoroacetic acid is removed by evaporation *in vacuo*, with subsequent trituration with diethyl ether affording the crude peptide. Any scavengers present are removed by a simple extraction procedure which on lyophilisation of the aqueous phase affords the crude peptide free of scavengers. Reagents for peptide synthesis are generally available from Calbiochem-Novabiochem (UK) Ltd, Nottingham NG7 2QJ, UK. Purification may be effected by any one, or a combination of, techniques such as size exclusion chromatography, ion-exchange chromatography and (principally) reverse-phase high performance liquid chromatography. Analysis of peptides may be carried out using thin layer chromatography, reverse-phase high performance liquid

30

chromatography, amino-acid analysis after acid hydrolysis and by fast atom bombardment (FAB) mass spectrometric analysis.

5   It will be appreciated that peptidomimetic compounds may also be useful. Thus, by "polypeptide" or "peptide" we include not only molecules in which amino acid residues are joined by peptide (-CO-NH-) linkages but also molecules in which the peptide bond is reversed. Such retro-inverso peptidomimetics may be made using methods known in the art, for example such as those described in Mézière *et al* (1997) *J. Immunol.* <u>159</u>, 3230-3237, incorporated herein by

10   reference. This approach involves making pseudopeptides containing changes involving the backbone, and not the orientation of side chains. Mézière *et al* (1997) show that, at least for MHC class II and T helper cell responses, these pseudopeptides are useful. Retro-inverse peptides, which contain NH-CO bonds instead of CO-NH peptide bonds, are much more resistant to proteolysis.

15

Similarly, the peptide bond may be dispensed with altogether provided that an appropriate linker moiety which retains the spacing between the Cα atoms of the amino acid residues is used; it is particularly preferred if the linker moiety has substantially the same charge distribution and substantially the same planarity as

20   a peptide bond.

It will be appreciated that the peptide may conveniently be blocked at its N- or C-terminus so as to help reduce susceptibility to exoproteolytic digestion.

25   Thus, it will be appreciated that the interacting polypeptide, for example which comprises the amino acid sequence PP(T/N)K may be a peptidomimetic compound, as described above.

31

A further aspect of the invention provides a molecule comprising an interacting polypeptide of the invention and a further portion, wherein the said molecule is not full-length *Xenopus* FAST1 or human FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer or *Xenopus* Bix2 (and may preferably not be Zebrafish FAST1 or Zebrafish Mixer). It is preferred that the said further portion confers a desirable feature on the said molecule; for example, the portion may useful in detecting or isolating the molecule, or promoting cellular uptake of the molecule or the interacting polypeptide. The portion may be, for example, a biotin moiety, a radioactive moiety, a fluorescent moiety, for example a small fluorophore or a green fluorescent protein (GFP) fluorophore, as well known to those skilled in the art. The moiety may be an immunogenic tag, for example a Myc tag, as known to those skilled in the art or may be a lipophilic molecule or polypeptide domain that is capable of promoting cellular uptake of the molecule or the interacting polypeptide, as known to those skilled in the art, for example as characterised for a *Drosophila* polypeptide. Thus, the moiety may derivable from the *Antennapedia* helix 3 (Derossi *et al* (1998) *Trends Cell Biol* **8**, 84-87).

A particularly preferred molecule of the invention is Biotin.Aminohexanoicacid-RQIKIWFQNRRMKWKKLLMDFNNFPPNKTITPDMNVRIPPI, discussed in Example 1. The first 16 amino acids are from the helix 3 of *Antennapedia* which allows internalization of these peptides into live cells (Derossi *et al* 1998); the last 25 amino acids are codons 283-307 of Mixer.

Further preferred molecules of the invention are the following, discussed in Example 2:

Mixer SIM peptide

32

Biotin.Aminohexanoicacid-

<u>RQIKIWFQNRRMKWKK</u>LLMDFNNFPPNKTITPDMNVRIPPI

Mixer SIM mutant peptide (not an interacting polypeptide)

Biotin.Aminohexanoicacid-

5   <u>RQIKIWFQNRRMKWKK</u>LLMDFNNFAANKTITPDMNVRIPPI

XFast-3 SIM peptide

5-FAM-AMINOHEXANOICACID-

<u>RQIKIWFQNRRMKWKK</u>PEVKNAPKDFPPNKTVFDIPVYTGHPGFLA

XFast-3 mutant SIM peptide (not an interacting polypeptide)

10   5-FAM-AMINOHEXANOICACID-

<u>RQIKIWFQNRRMKWKK</u>PEVKNAPKDFAAAKTVFDIPVYTGHPGFLA

where 5-FAM is 5-carboxyfluorescein (C1359 from Molecular Probes).

The Antennapedia third helix is underlined.

15

A further aspect of the invention provides a nucleic acid (or polynucleotide) encoding or capable of expressing an interacting polypeptide or polypeptide containing PP(T/N)K of the invention. A still further aspect of the invention provides a nucleic acid complementary to a nucleic acid encoding or capable of

20   expressing a polypeptide of the invention. Methods of preparing or isolating such a nucleic acid are well known to those skilled in the art.

The following methods of isolating a nucleic acid encoding an interacting polypeptide or polypeptide containing PP(T/N)K of the invention are given for

25   purposes of illustration and are not considered to be exhaustive.

33
The polypeptide may be cleaved, for example using trypsin, cyanogen bromide, V8 protease formic acid, or another specific cleavage reagent. The digest may be chromatographed on a Vydac C18 column or subjected to SDS-PAGE to resolve the peptides. The N-terminal sequence of the peptides may then be
5 determined using standard methods.

The sequences are used to isolate a nucleic acid encoding the peptide sequences using standard PCR-based strategies. Degenerate oligonucleotide mixtures, each comprising a mixture of all possible sequences encoding a part of the peptide
10 sequences, are designed and used as PCR primers or probes for hybridisation analysis of PCR products after Southern blotting. mRNA prepared from cells in which the polypeptide may be expressed is used as the template for reverse transcriptase, to prepare cDNA, which is then used as the template for the PCR reactions.
15
Positive PCR fragments are subcloned and used to screen cDNA libraries to isolate a full length clone for the polypeptide.

Alternatively, the sequences of initial subcloned PCR fragments may be
20 determined, and the sequence may then be extended by known PCR-based techniques to obtain a full length sequence.

Alternatively, the initial PCR sequence may be used to screen electronic databases of expressed sequence tags (ESTs) or other known sequences. By this
25 means, related sequences may be identified which may be useful in isolating a full length sequence using the two approaches described above.

34

Sequences are determined using the Sanger dideoxy method. The encoded amino acid sequences may be deduced by routine methods.

Techniques used are essentially as described in Sambrook *et al* (1989) Molecular cloning, a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Alternatively, antibodies may be raised against the polypeptide.

The antibodies are used to screen a λgt11 expression library made from cDNA copied from mRNA from cells in which the polypeptide may be expressed.

Positive clones are identified and the insert sequenced by the Sanger method as mentioned above. The encoded amino acid sequence may be deduced by routine methods.

It will be appreciated that it may be desirable to express the polypeptide encoded by the isolated nucleic acid in order to determine that the polypeptide has the expected properties, for example that it is capable of interacting with a Smad polypeptide, for example Smad2 or Smad3.

The invention also includes a polynucleotide comprising a fragment of the recombinant polynucleotide of the second aspect of the invention. Preferably, the polynucleotide comprises a fragment which is at least 10 nucleotides in length, more preferably at least 14 nucleotides in length and still more preferably at least 18 nucleotides in length. Such polynucleotides are useful as PCR primers.

35

The polynucleotide or recombinant polynucleotide may be DNA or RNA, preferably DNA. The polynucleotide may or may not contain introns in the coding sequence; preferably the polynucleotide is a cDNA.

5

A "variation" of the polynucleotide includes one which is (i) usable to produce a protein or a fragment thereof which is in turn usable to prepare antibodies which specifically bind to the protein encoded by the said polynucleotide or (ii) an antisense sequence corresponding to the gene or to a variation of type (i) as just

10   defined. For example, different codons can be substituted which code for the same amino acid(s) as the original codons. Alternatively, the substitute codons may code for a different amino acid that will not affect the activity or immunogenicity of the protein or which may improve or otherwise modulate its activity or immunogenicity. For example, site-directed mutagenesis or other

15   techniques can be employed to create single or multiple mutations, such as replacements, insertions, deletions, and transpositions, as described in Botstein and Shortle, "Strategies and Applications of *In Vitro* Mutagenesis," *Science*, **229**: 193-210 (1985), which is incorporated herein by reference. Since such modified polynucleotides can be obtained by the application of known techniques to the

20   teachings contained herein, such modified polynucleotides are within the scope of the claimed invention.

Moreover, it will be recognised by those skilled in the art that the polynucleotide sequence (or fragments thereof) of the invention can be used to obtain other

25   polynucleotide sequences that hybridise with it under conditions of high stringency. Such polynucleotides includes any genomic DNA. Accordingly, the polynucleotide of the invention includes polynucleotide that shows at least 55 per

36

cent, preferably 60 per cent, and more preferably at least 70 per cent and most preferably at least 90 per cent homology with the polynucleotide identified in the method of the invention, provided that such homologous polynucleotide encodes a polypeptide which is usable in at least some of the methods described below or
5   is otherwise useful.

Per cent homology can be determined by, for example, the GAP program of the University of Wisconsin Genetic Computer Group.

10   DNA-DNA, DNA-RNA and RNA-RNA hybridisation may be performed in aqueous solution containing between 0.1XSSC and 6XSSC and at temperatures of between 55°C and 70°C. It is well known in the art that the higher the temperature or the lower the SSC concentration the more stringent the hybridisation conditions. By "high stringency" we mean 2XSSC and 65°C.
15   1XSSC is 0.15M NaCl/0.015M sodium citrate. Polynucleotides which hybridise at high stringency are included within the scope of the claimed invention.

"Variations" of the polynucleotide also include polynucleotide in which relatively short stretches (for example 20 to 50 nucleotides) have a high degree
20   of homology (at least 80% and preferably at least 90 or 95%) with equivalent stretches of the polynucleotide of the invention even though the overall homology between the two polynucleotides may be much less. This is because important active or binding sites may be shared even when the general architecture of the protein is different.
25

A further aspect of the invention provides a replicable vector comprising a recombinant polynucleotide encoding an interacting polypeptide or a polypeptide

37

containing PP(T/N)K of the invention. It will be appreciated that the said recombinant polynucleotide may encode an interacting polypeptide or polypeptide containing PP(T/N)K of the invention that is a fusion of an interacting polypeptide or polypeptide containing PP(T/N)K.

A variety of methods have been developed to operably link polynucleotides, especially DNA, to vectors for example *via* complementary cohesive termini. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

Synthetic linkers containing one or more restriction sites provide an alternative method of joining the DNA segment to vectors. The DNA segment, generated by endonuclease restriction digestion as described earlier, is treated with bacteriophage T4 DNA polymerase or *E. coli* DNA polymerase I, enzymes that remove protruding, 3'-single-stranded termini with their 3'-5'-exonucleolytic activities, and fill in recessed 3'-ends with their polymerizing activities.

The combination of these activities therefore generates blunt-ended DNA segments. The blunt-ended segments are then incubated with a large molar excess of linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteriophage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying polymeric linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction enzyme and ligated to an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the DNA

38

segment.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including International
5   Biotechnologies Inc, New Haven, CN, USA.

A desirable way to modify the DNA encoding the polypeptide of the invention is to use the polymerase chain reaction as disclosed by Saiki *et al* (1988) *Science* 239, 487-491. This method may be used for introducing the DNA into a suitable
10   vector, for example by engineering in suitable restriction sites, or it may be used to modify the DNA in other useful ways as is known in the art.

In this method the DNA to be enzymatically amplified is flanked by two specific primers which themselves become incorporated into the amplified DNA. The
15   said specific primers may contain restriction endonuclease recognition sites which can be used for cloning into expression vectors using methods known in the art.

The DNA (or in the case of retroviral vectors, RNA) is then expressed in a
20   suitable host to produce a polypeptide comprising the compound of the invention. Thus, the DNA encoding the polypeptide of the invention may be used in accordance with known techniques, appropriately modified in view of the teachings contained herein, to construct an expression vector, which is then used to transform an appropriate host cell for the expression and production of the
25   polypeptide of the invention. Such techniques include those disclosed in US Patent Nos. 4,440,859 issued 3 April 1984 to Rutter *et al*, 4,530,901 issued 23 July 1985 to Weissman, 4,582,800 issued 15 April 1986 to Crowl, 4,677,063

issued 30 June 1987 to Mark *et al*, 4,678,751 issued 7 July 1987 to Goeddel, 4,704,362 issued 3 November 1987 to Itakura *et al*, 4,710,463 issued 1 December 1987 to Murray, 4,757,006 issued 12 July 1988 to Toole, Jr. *et al*, 4,766,075 issued 23 August 1988 to Goeddel *et al* and 4,810,648 issued 7 March

5   1989 to Stalker, all of which are incorporated herein by reference.

The DNA (or in the case of retroviral vectors, RNA) encoding the polypeptide constituting the compound of the invention may be joined to a wide variety of other DNA sequences for introduction into an appropriate host. The companion

10   DNA will depend upon the nature of the host, the manner of the introduction of the DNA into the host, and whether episomal maintenance or integration is desired.

Generally, the DNA is inserted into an expression vector, such as a plasmid, in

15   proper orientation and correct reading frame for expression. If necessary, the DNA may be linked to the appropriate transcriptional and translational regulatory control nucleotide sequences recognised by the desired host, although such controls are generally available in the expression vector. The vector is then introduced into the host through standard techniques. Generally, not all of the

20   hosts will be transformed by the vector. Therefore, it will be necessary to select for transformed host cells. One selection technique involves incorporating into the expression vector a DNA sequence, with any necessary control elements, that codes for a selectable trait in the transformed cell, such as antibiotic resistance. Alternatively, the gene for such selectable trait can be on another

25   vector, which is used to co-transform the desired host cell.

Host cells that have been transformed by the recombinant DNA of the invention

40

are then cultured for a sufficient time and under appropriate conditions known to those skilled in the art in view of the teachings disclosed herein to permit the expression of the polypeptide, which can then be recovered.

5   Many expression systems are known, including bacteria (for example *E. coli* and *Bacillus subtilis*), yeasts (for example *Saccharomyces cerevisiae*), filamentous fungi (for example *Aspergillus*), plant cells, animal cells and insect cells.

The vectors include a prokaryotic replicon, such as the ColE1 *ori*, for
10   propagation in a prokaryote, even if the vector is to be used for expression in other, non-prokaryotic, cell types. The vectors can also include an appropriate promoter such as a prokaryotic promoter capable of directing the expression (transcription and translation) of the genes in a bacterial host cell, such as *E. coli*, transformed therewith.
15

A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with exemplary bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA
20   segment of the present invention.

Typical prokaryotic vector plasmids are pUC18, pUC19, pBR322 and pBR329 available from Biorad Laboratories, (Richmond, CA, USA) and p*Trc*99A and pKK223-3 available from Pharmacia, Piscataway, NJ, USA.
25

A typical mammalian cell vector plasmid is pSVL available from Pharmacia, Piscataway, NJ, USA. This vector uses the SV40 late promoter to drive

41

expression of cloned genes, the highest level of expression being found in T antigen-producing cells, such as COS-1 cells.

An example of an inducible mammalian expression vector is pMSG, also
5   available from Pharmacia.   This vector uses the glucocorticoid-inducible promoter of the mouse mammary tumour virus long terminal repeat to drive expression of the cloned gene.

Useful yeast plasmid vectors are pRS403-406 and pRS413-416 and are generally
10   available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Plasmids pRS403, pRS404, pRS405 and pRS406 are Yeast Integrating plasmids (YIps) and incorporate the yeast selectable markers *HIS3*, *TRP1*, *LEU2* and *URA3*. Plasmids pRS413-416 are Yeast Centromere plasmids (YCps).

15   The present invention also relates to a host cell transformed with a polynucleotide vector construct of the present invention.   The host cell can be either prokaryotic or eukaryotic.   Bacterial cells are preferred prokaryotic host cells and typically are a strain of *E. coli* such as, for example, the *E. coli* strains DH5 available from Bethesda Research Laboratories Inc., Bethesda, MD, USA,
20   and RR1 available from the American Type Culture Collection (ATCC) of Rockville, MD, USA (No ATCC 31343).   Preferred eukaryotic host cells include yeast, insect and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human fibroblastic and kidney cell lines. Yeast host cells include YPH499, YPH500 and YPH501 which are generally
25   available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Preferred mammalian host cells include Chinese hamster ovary (CHO) cells available from the ATCC as CCL61, NIH Swiss mouse embryo cells NIH/3T3

42

available from the ATCC as CRL 1658, monkey kidney-derived COS-1 cells available from the ATCC as CRL 1650 and 293 cells which are human embryonic kidney cells. Preferred insect cells are Sf9 cells which can be transfected with baculovirus expression vectors.

5

Transformation of appropriate cell hosts with a DNA construct of the present invention is accomplished by well known methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen *et al* (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110 and

10   Sambrook *et al* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Transformation of yeast cells is described in Sherman *et al* (1986) *Methods In Yeast Genetics, A Laboratory Manual*, Cold Spring Harbor, NY. The method of Beggs (1978) *Nature* **275**, 104-109 is also useful. With regard to vertebrate cells, reagents useful in

15   transfecting such cells, for example calcium phosphate and DEAE-dextran or liposome formulations, are available from Stratagene Cloning Systems, or Life Technologies Inc., Gaithersburg, MD 20877, USA.

Electroporation is also useful for transforming and/or transfecting cells and is

20   well known in the art for transforming yeast cell, bacterial cells, insect cells and vertebrate cells.

For example, many bacterial species may be transformed by the methods described in Luchansky *et al* (1988) *Mol. Microbiol.* **2**, 637-646 incorporated

25   herein by reference. The greatest number of transformants is consistently recovered following electroporation of the DNA-cell mixture suspended in 2.5X PEB using 6250V per cm at 25μFD.

43

Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) *Methods Enzymol.* 194, 182.

5   Successfully transformed cells, ie cells that contain a DNA construct of the present invention, can be identified by well known techniques. For example, cells resulting from the introduction of an expression construct of the present invention can be grown to produce the polypeptide of the invention. Cells can be harvested and lysed and their DNA content examined for the presence of the

10   DNA using a method such as that described by Southern (1975) *J. Mol. Biol.* **98**, 503 or Berent *et al* (1985) *Biotech.* **3**, 208. Alternatively, the presence of the protein in the supernatant can be detected using antibodies as described below.

In addition to directly assaying for the presence of recombinant DNA, successful

15   transformation can be confirmed by well known immunological methods when the recombinant DNA is capable of directing the expression of the protein. For example, cells successfully transformed with an expression vector produce proteins displaying appropriate antigenicity. Samples of cells suspected of being transformed are harvested and assayed for the protein using suitable antibodies.

20

Thus, in addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, preferably a monoclonal (clonally homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium.

25

A further aspect of the invention provides a method of making a polypeptide of the invention the method comprising culturing a host cell comprising a

44

recombinant polynucleotide or a replicable vector which encodes said polypeptide, and isolating said polypeptide from said host cell. Methods of cultivating host cells and isolating recombinant proteins are well known in the art.

5

A further aspect of the invention provides an antibody capable of reacting with a polypeptide of the invention, in particular an antibody capable of reacting with an epitope comprising the amino acid sequence PP(T/N)K. Antibodies reactive towards the said polypeptide of the invention may be made by methods well

10   known in the art. In particular, the antibodies may be polyclonal or monoclonal.

Suitable monoclonal antibodies may be prepared by known techniques, for example those disclosed in "Monoclonal Antibodies: A manual of techniques", H Zola (CRC Press, 1988) and in "Monoclonal Hybridoma Antibodies:

15   Techniques and applications", J G R Hurrell (CRC Press, 1982), both of which are incorporated herein by reference. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations useful in the methods claimed in this invention. Techniques for preparing antibodies are well known to those skilled in the art,

20   for example as described in Harlow, ED & Lane, D "Antibodies: a laboratory manual" (1988) New York Cold Spring Harbor Laboratory.

Polyclonal antibodoes may be prepared using methods well known in the art. In the case of both monoclonal and polyclonal antibodies, it is useful to use as

25   immunogen any suitable polypeptide containing a SIM, for example containing the PP(T/N)K motif. In particular with respect to the production of polyclonal

45

antibodies it is useful to use polypeptides of between 10 and 30 amino acid residues containing a SIM, for example containing the PP(T/N)K motif.

5   In a preferred embodiment of the invention, an antibody of the invention is capable of preventing or disrupting the interaction between a Smad polypeptide and a polypeptide comprising a SIM, for example comprising the amino acid sequence PP(T/N)K.

It will be appreciated that other antibody-like molecules may be useful in the

10   practice of the invention including, for example, antibody fragments or derivatives which retain their antigen-binding sites, synthetic antibody-like molecules such as single-chain Fv fragments (ScFv) and domain antibodies (dAbs), and other molecules with antibody-like antigen binding motifs. Such antibody-like molecules are included by the term antibody as used below.

15

A further aspect of the invention provides a method of disrupting or preventing the interaction between a Smad polypeptide and a polypeptide (target polypeptide) that is (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with the said

20   Smad polypeptide, the interaction requiring α-helix2 of the said Smad polypeptide, the method comprising exposing the Smad polypeptide to an interacting polypeptide of the invention or an antibody of the invention. Alternatively, the Smad polypeptide may be exposed to a compound of the invention, as described below. It will be appreciated that the said polypeptide

25   capable of interacting with the said Smad polypeptide may interact with α-helix 2 of the said Smad polypeptide; alternatively, the interaction may require α-helix 2 but contact between the said polypeptide capable of interacting with the said

46

Smad polypeptide and the said Smad polypeptide may occur at site in the said Smad polypeptide that is not part of α-helix 2.

A further aspect of the invention provides a method of disrupting or preventing
5   the interaction between a Smad polypeptide and a polypeptide (target polypeptide) which target polypeptide comprises a SIM, for example comprises the amino acid sequence PP(T/N)K, the method comprising exposing the Smad polypeptide to an interacting polypeptide of the invention or an antibody of the invention. Alternatively, the Smad polypeptide may be exposed to a compound
10   of the invention, as described below.

Preferences for the SIM and for the Smad polypeptide are as set out in relation to earlier aspects of the invention. It is particularly preferred that the Smad polypeptide is a naturally occurring Smad polypeptide, for example Smad2 or
15   Smad3 or naturally occurring allelic variants thereof. It is still more preferred that the Smad polypeptide is a human Smad polypeptide, for example human Smad2 or human Smad3.

It is preferred that the antibody of the invention is capable of reacting with an
20   epitope comprising the amino acid sequence PP(T/N)K.

The target polypeptide may be an interacting polypeptide of the invention, for example FAST3. It is preferred that the target polypeptide comprises a SIM, for example comprises the amino acid sequence PP(T/N)K. The target polypeptide
25   may be FAST1, FAST2, Mixer, Milk or Bix 2 or 3 or a fragment, variant, derivative or fusion thereof. It is preferred that the target polypeptide is a naturally occurring polypeptide or a fusion thereof.

The interaction between the Smad polypeptide and the target polypeptide and its disruption or prevention may be measured by any method of detecting/measuring a protein/protein interaction, as discussed further below and in Example 1.

5   Suitable methods include yeast two-hybrid interactions, co-purification, ELISA, co-immunoprecipitation methods and bandshift assays.

The methods may be performed *in vitro*, either in intact cells or tissues, with broken cell or tissue preparations or at least partially purified components.

10   Alternatively, they may be performed *in vivo*. The cells tissues or organisms in/on which the use or methods are performed may be transgenic. In particular they may be transgenic for the Smad interacting protein under consideration or for a further Smad interacting protein or Smad.

15   A further aspect of the invention provides a method of identifying a polypeptide (interacting polypeptide) that is capable of interacting with a Smad polypeptide, for example Smad2 or Smad3, comprising examining the sequence of a polypeptide and determining that the polypeptide comprises a SIM, for example comprises the amino acid sequence PP(T/N)K (or three out of four residues

20   thereof). It is believed that the amino acid sequence PP(T/N)K (or at three out of four residues thereof) is necessary and may be sufficient for interaction of a polypeptide with a Smad polypeptide, for example Smad2 or Smad3. Preferences for the Smad polypeptide are as given above.

25   The presence of further or other features of a SIM may be determined. It may be determined that an acidic amino acid residue is present at a position from 3 to 10, preferably 4 to 5 residues residues C-terminal of the amino acid sequence

48

corresponding to the PP(T/N)K motif (which may of course be PP(T/N)K) and/or a proline residue is present at a position from 5 to 20 residues C-terminal of the amino acid sequence corresponding to the PP(T/N)K motif; these residues may also promote the interaction between the said interacting polypeptide and

5   the Smad polypeptide. The acidic (negatively charged) amino acid residue is typically a glutamate or aspartate residue. It may further be determined that the said acidic residue (ie present at position about +3 to about +10, preferably +4 or +5) is immediately followed by a hydrophobic residue, for example M, V or I. The downstream proline and acidic, for example aspartate residues as

10   described above may also be important for binding. It may further be determined that the residue immediately after (ie C-terminal of) the amino acid sequence corresponding to the PP(T/N)K motif ie at position +1 is an S or T, which may be immediately followed by an I or V residue. It may further be determined that the residue immediately before (ie N-terminal of) the amino acid sequence

15   corresponding to the PP(T/N)K motif is a hydrophobic residue, for example F, M or V. It may further be determined that a proline residue is present at a position from 5 to 20 residues C-terminal of the amino acid sequence corresponding to the sequence motif PP(T/N)K.

20   It may further be determined that an acidic residue (for example glutamate or aspartate) immediately followed by a hydrophobic residue (for example F, Y, L) may be present at position starting about -20 to -2 relative to the amino acids corresponding to the PP(T/N)K sequence motif, preferably at -9 to -8 or -5 to -4 or -2 to -1 (ie immediately N-terminal of the PP(T/N)K sequence motif. It

25   may further be determined that a leucine residue may be present at position about -2 to -15, preferably about -5 to -10. The leucine residue may be the

49

hydrophobic residue that is immediately preceded by an acidic residue, as noted above.

5   The method may comprise determining that the polypeptide comprises at least 8, 9 or 10 of the specified residues (ie not residues designated by an X) of the amino acid sequence D/E-Hyd-$(X)_n$-P-P-(N/T)-K-(T/S)-(I/V)-$(X)_m$-(D/E)-(M/V/I)-$(X)_k$-P

wherein m= 0 to 7; k= 0 to 8 or 12; n = 0 to 15 or 18.

10

Should the amino acid sequence of the said interacting polypeptide or the nucleotide sequence encoding the said interacting polypeptide not be known, they may be determined by methods well known to those skilled in the art, for example PCR-based cloning methods, as indicated above. It may be desirable to

15   confirm that the interacting polypeptide identified by the method is capable of interacting with a Smad polypeptide, for example Smad2 or Smad3, using methods of detecting or measuring protein/protein interactions, as described above, for example using the interacting polypeptide expressed as described above.

20

The interacting polypeptide may also be useful in a screening assay for identifying a drug like compound that may inhibit the interaction between Smad2 or Smad3 and a polypeptide that interacts with Smad2 or Smad3 *in vivo*, for example a homologue of Milk, Mixer, other Mix family members, FAST1 and

25   FAST2. It will be appreciated that the polypeptide may only interact with Smad2 or Smad3 when the Smad2 or Smad3 is in an activated state, for example following activation and/or phosphorylation as a consequence of TGFβ

50

superfamily receptor activation, or wherein the N-terminal domain is not present or is truncated. It will be appreciated that the Smad2 or Smad3 may further interact with Smad4. It will be further appreciated that the Smad2 or Smad3 may interact or form a complex with more than one polypeptide that is not

5 Smad4; for example, Smad2 or Smad3 may form a complex with Mixer, Milk and Smad4. Mixer and Milk may form a heterodimer.

A further aspect of the invention thus provides a method of identifying a compound capable of disrupting or preventing the interaction between a Smad

10 polypeptide and a polypeptide (target polypeptide) that is (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with a Smad polypeptide, the interaction requiring α-helix2 of the said Smad polypeptide and/or (3) a polypeptide comprising the amino acid sequence PP(T/N)K, the method comprising

15 measuring the ability of the compound to disrupt or prevent the interaction between the Smad polypeptide and an interacting polypeptide of the invention.

The interaction between the Smad polypeptide and the interacting polypeptide and its disruption or prevention may be measured by any method of

20 detecting/measuring a protein/protein interaction, as discussed in Example 1. Suitable methods include yeast two-hybrid interactions, co-purification, ELISA, co-immunoprecipitation methods and bandshift assays. Further suitable methods may include Scintillation Proximity Assays, as well known to those skilled in the art. Examples of suitable methods may include bandshift assays looking for

25 disruption of the endogenous FAST/Smad2/Smad4 ARF complex or disruption of the Mixer/GSTSmad2C interaction, as described in Example 1 and transcription assays in tissue culture cells in which expression of a reporter gene

51

driven by a promoter with a binding site for (for example) Mixer is measured following treatment of the cells with TGFβ. Disruption or prevention of TGFβ-dependent transcription in the presence of the compound may be detected. The cells may be transiently transfected or may be a stable cell line capable of

5   expressing Mixer (or other appropriate transcription factor) with an integrated reporter gene, as described in Example 2. The reporter gene may express luciferase or a green fluorescent protein (GFP) or secreted alkaline phosphatase (SEAP) or CAT, as well known to those skilled in the art.

10   It will be appreciated that chip screening methods may be used. For example, arrays of cDNAs or oligonucleotides may be used in assessing expression of endogenous genes that are modulated by TGFβ and therefore for assessing effects of compounds on such expression.

15   A further aspect of the invention therefore provides a cell comprising 1) a recombinant polynucleotide suitable for expressing a transcription factor that is capable of interacting with a Smad polypeptide and 2) a recombinant polynucleotide comprising a reporter gene driven by a promoter with a binding site for the said transcription factor. A further aspect of the invention provides

20   a stably-transformed cell line cell comprising a reporter gene driven by a promoter with a binding site for an activated Smad, wherein the Smad is activated in the cell by exposure of the cell to TGFβ. The reporter gene may express luciferase or CAT or SEAP or a green fluorescent protein (GFP). A further aspect of the invention provides a method of identifying a compound

25   capable of modulating TGFβ-dependent transcription wherein the effect of the compound on expression of the reporter gene in a cell of the invention is measured, following treatment of the cell with TGFβ.

52

A further aspect of the invention provides a method of identifying a compound capable of modulating TGFβ-dependent transcription wherein the effect of the compound on TGFβ-signalling- dependent invasive behaviour of a stably-transformed cell line cell, for example in collagen gels, is measured and a compound that reduces invasive behaviour is selected. The stably-transformed cell line is preferably a MDCK cell line that is capable of expressing recombinant active Raf-1, as described in Example 2.

10   The methods of the invention may be performed *in vitro*, either in intact cells or tissues, with broken cell or tissue preparations or at least partially purified components. Alternatively, they may be performed *in vivo*. The cells tissues or organisms in/on which the use or methods are performed may be transgenic. In particular they may be transgenic for the Smad interacting protein under

15   consideration or for a further Smad interacting protein or Smad. Thus, a transgenic animal, for example a transgenic rodent, for example mouse or rat, amphibian, for example *Xenopus*, or insect, for example *Drosophila*, transgenic for the Smad interacting protein under consideration or for a further Smad interacting protein or Smad may be useful, for example in the screening methods

20   of the invention.

It will be appreciated that screening assays which are capable of high throughput operation will be particularly preferred. Examples may include cell based assays, for example as described in Chen *et al* (1997) and protein-protein

25   binding assays. An SPA-based (Scintillation Proximity Assay; Amersham International) system may be used. For example, beads comprising scintillant and a Smad polypeptide, for example Smad2 or a fragment (for example the

53

MH2 domain) may be prepared. The beads may be mixed with a sample comprising the interacting polypeptide into which a radioactive label has been incorporated and with the test compound. Conveniently this is done in a 96-well format. The plate is then counted using a suitable scintillation counter, using known parameters for the particular radioactive label in an SPA assay. Only the radioactive label that is in proximity to the scintillant, ie only that bound to the interacting polypeptide that is bound to the Smad polypeptide anchored on the beads, is detected. Variants of such an assay, for example in which the Smad polypeptide is immobilised on the scintillant beads *via* binding to an antibody or antibody fragment, may also be used.

Other methods of detecting polypeptide/polypeptide interactions include ultrafiltration with ion spray mass spectroscopy/HPLC methods or other physical and analytical methods. Fluorescence Energy Resonance Transfer (FRET) methods, for example, well known to those skilled in the art, may be used, in which binding of two fluorescent labeled entities may be measured by measuring the interaction of the fluorescent labels when in close proximity to each other.

The compound may be a drug-like compound or lead compound for the development of a drug-like compound for each of the above methods of identifying a compound. It will be appreciated that the said methods may be useful as screening assays in the development of pharmaceutical compounds or drugs, as well known to those skilled in the art.

The term "drug-like compound" is well known to those skilled in the art, and may include the meaning of a compound that has characteristics that may make it suitable for use in medicine, for example as the active ingredient in a

54

medicament. Thus, for example, a drug-like compound may be a molecule that may be synthesised by the techniques of organic chemistry, less preferably by techniques of molecular biology or biochemistry, and is preferably a small molecule, which may be of less than 5000 daltons molecular weight. A drug-

5   like compound may additionally exhibit features of selective interaction with a particular protein or proteins and be bioavailable and/or able to penetrate cellular membranes, but it will be appreciated that these features are not essential.

The term "lead compound" is similarly well known to those skilled in the art,
10   and may include the meaning that the compound, whilst not itself suitable for use as a drug (for example because it is only weakly potent against its intended target, non-selective in its action, unstable, difficult to synthesise or has poor bioavailability) may provide a starting-point for the design of other compounds that may have more desirable characteristics.

15

It will be appreciated that the compound may be a polypeptide that is capable of competing with the interacting polypeptide of the invention for binding to the Smad polypeptide, and may be (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting
20   with a Smad polypeptide, the interaction requiring α-helix2 of the said Smad polypeptide and/or (3) a polypeptide comprising a SIM, for example comprising the amino acid sequence PP(T/N)K. Thus, it will be appreciated that a screening method as described above may be useful in identifying polypeptides that may interact with the Smad polypeptide.

25

Methods that may be useful in identifying polypeptides that may interact with the Smad polypeptide include yeast-2-hybrid, co-immunoprecipitation, ELISA, GST-

55

pulldown, bandshift and transcription assays. Transcription assays may be performed *in vivo* or *in vitro*. For example, tissue culture cells may be used which comprise a reporter construct in which expression of the reporter gene is controlled a promoter comprising a binding site(s) for the putative interacting

5   polypeptide. The effect of treating the cells with TGFβ on expression of the reporter gene may then be measured; TGFβ-dependent expression of the reporter gene may indicate that the putative interacting polypeptide is capable of being regulated by TGFβ and therefore may interact with the said Smad polypeptide. It will be appreciated that a transcription assay may be performed in a transgenic

10   animal, for example a transgenic *Drosophila* or *Xenopus*.

A further aspect of the invention is a kit of parts useful in carrying out a method, for example a screening method, of the invention. Such a kit may comprise a Smad polypeptide, for example Smad2 or Smad3 or a fragment either therof and

15   an interacting polypeptide, for example a polypeptide corresponding to amino acids 283 to 307 of Mixer.

A further aspect of the invention provides a compound identified by or identifiable by the screening method of the invention.

20

It will be appreciated that such a compound may be an inhibitor of the formation or stability of a complex of the Smad polypeptide used in the screen, for example Smad2 or Smad3, with interacting polypeptide(s), for example Smad4 and a transcription factor, for example FAST1, FAST2, Mixer, Milk or Bix2/3, and

25   therefore ultimately of the activity of that complex, for example in promoting the transcription from a promoter to which the complex binds. The intention of the screen may be to identify compounds that act as modulators, for example

56

inhibitors or promoters, preferably inhibitors of the activity of the complex, even if the screen makes use of a binding assay rather than an activity (for example transcriptional activity or DNA binding) assay. It will be appreciated that the inhibitory action of a compound found to bind the Smad or Smad interacting

5   polypeptide may be confirmed by performing an assay of, for example, transcriptional or DNA binding activity in the presence of the compound.

A further aspect of the invention provides a compound identified by or identifiable by the screening method of the invention for use in medicine. A still

10   further aspect of the invention provides an interacting polypeptide or polypeptide containing PP(T/N)K or molecule of the invention or nucleic acid of the invention or antibody of the invention for use in medicine.

The compound, interacting polypeptide, polypeptide containing PP(T/N)K,

15   molecule, nucleic acid or antibody of the invention is suitably packaged and presented for use in medicine.

The aforementioned interacting polypeptide or molecule of the invention or nucleic acid of the invention or antibody of the invention or a formulation

20   thereof may be administered by any conventional method including oral and parenteral (e.g. subcutaneous or intramuscular) injection. The treatment may consist of a single dose or a plurality of doses over a period of time.

Whilst it is possible for an interacting polypeptide or molecule of the invention

25   or nucleic acid of the invention or antibody of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers. The carrier(s) must be "acceptable" in the sense of being compatible with the compound of the invention and not

57

deleterious to the recipients thereof. Typically, the carriers will be water or saline which will be sterile and pyrogen free.

5   Thus, the invention also provides pharmaceutical compositions comprising the interacting polypeptide or molecule of the invention or nucleic acid of the invention or antibody of the invention and a pharmaceutically acceptable carrier.

As indicated above, the nucleic acid of the invention may be an antisense oligonucleotide, for example an antisense oligonucleotide directed against a

10   nucleic acid encoding an interacting polypeptide of the invention, which may be a transcription factor comprising a SIM, for example comprising the amino acid sequence PP(N/T)K. It is preferred that the antisense oligonucleotide is directed against a nucleic acid encoding a human transcription factor.

15   Antisense oligonucleotides are single-stranded nucleic acid, which can specifically bind to a complementary nucleic acid sequence. By binding to the appropriate target sequence, an RNA-RNA, a DNA-DNA, or RNA-DNA duplex is formed. These nucleic acids are often termed "antisense" because they are complementary to the sense or coding strand of the gene. Recently, formation of

20   a triple helix has proven possible where the oligonucleotide is bound to a DNA duplex. It was found that oligonucleotides could recognise sequences in the major groove of the DNA double helix. A triple helix was formed thereby. This suggests that it is possible to synthesise a sequence-specific molecules which specifically bind double-stranded DNA via recognition of major groove

25   hydrogen binding sites.

58

By binding to the target nucleic acid, the above oligonucleotides can inhibit the function of the target nucleic acid. This could, for example, be a result of blocking the transcription, processing, poly(A)addition, replication, translation, or promoting inhibitory mechanisms of the cells, such as promoting RNA

5   degradations.

Antisense oligonucleotides are prepared in the laboratory and then introduced into cells, for example by microinjection or uptake from the cell culture medium into the cells, or they are expressed in cells after transfection with plasmids or

10   retroviruses or other vectors carrying an antisense gene. Antisense oligonucleotides were first discovered to inhibit viral replication or expression in cell culture for Rous sarcoma virus, vesicular stomatitis virus, herpes simplex virus type 1, simian virus and influenza virus. Since then, inhibition of mRNA translation by antisense oligonucleotides has been studied extensively in cell-free

15   systems including rabbit reticulocyte lysates and wheat germ extracts. Inhibition of viral function by antisense oligonucleotides has been demonstrated *in vitro* using oligonucleotides which were complementary to the AIDS HIV retrovirus RNA (Goodchild, J. 1988 "Inhibition of Human Immunodeficiency Virus Replication by Antisense Oligodeoxynucleotides", *Proc. Natl. Acad. Sci. (USA)*

20   85(15), 5507-11). The Goodchild study showed that oligonucleotides that were most effective were complementary to the poly(A) signal; also effective were those targeted at the 5' end of the RNA, particularly the cap and 5' untranslated region, next to the primer binding site and at the primer binding site. The cap, 5' untranslated region, and poly(A) signal lie within the sequence repeated at the

25   ends of retrovirus RNA (R region) and the oligonucleotides complementary to these may bind twice to the RNA.

59

Oligonucleotides are subject to being degraded or inactivated by cellular endogenous nucleases. To counter this problem, it is possible to use modified oligonucleotides, eg having altered internucleotide linkages, in which the naturally occurring phosphodiester linkages have been replaced with another linkage. For

5 example, Agrawal *et al* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7079-7083 showed increased inhibition in tissue culture of HIV-1 using oligonucleotide phosphoramidates and phosphorothioates. Sarin *et al* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7448-7451 demonstrated increased inhibition of HIV-1 using oligonucleotide methylphosphonates. Agrawal *et al* (1989) *Proc. Natl. Acad. Sci.*

10 *USA* **86**, 7790-7794 showed inhibition of HIV-1 replication in both early-infected and chronically infected cell cultures, using nucleotide sequence-specific oligonucleotide phosphorothioates. Leither *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3430-3434 report inhibition in tissue culture of influenza virus replication by oligonucleotide phosphorothioates.

15

Oligonucleotides having artificial linkages have been shown to be resistant to degradation *in vivo*. For example, Shaw *et al* (1991) in *Nucleic Acids Res.* **19**, 747-750, report that otherwise unmodified oligonucleotides become more resistant to nucleases *in vivo* when they are blocked at the 3′ end by certain capping

20 structures and that uncapped oligonucleotide phosphorothioates are not degraded *in vivo*.

A detailed description of the H-phosphonate approach to synthesising oligonucleoside phosphorothioates is provided in Agrawal and Tang (1990)

25 *Tetrahedron Letters* **31**, 7541-7544, the teachings of which are hereby incorporated herein by reference. Syntheses of oligonucleoside methylphosphonates, phosphorodithioates, phosphoramidates, phosphate esters,

60

bridged phosphoramidates and bridge phosphorothioates are known in the art. See, for example, Agrawal and Goodchild (1987) *Tetrahedron Letters* **28**, 3539; Nielsen *et al* (1988) *Tetrahedron Letters* **29**, 2911; Jager *et al* (1988) *Biochemistry* **27**, 7237; Uznanski *et al* (1987) *Tetrahedron Letters* **28**, 3401; Bannwarth (1988)

5   *Helv. Chim. Acta.* **71**, 1517; Crosstick and Vyle (1989) *Tetrahedron Letters* **30**, 4693; Agrawal *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1401-1405, the teachings of which are incorporated herein by reference. Other methods for synthesis or production also are possible. In a preferred embodiment the oligonucleotide is a deoxyribonucleic acid (DNA), although ribonucleic acid

10   (RNA) sequences may also be synthesised and applied.

The oligonucleotides useful in the invention preferably are designed to resist degradation by endogenous nucleolytic enzymes. *In vivo* degradation of oligonucleotides produces oligonucleotide breakdown products of reduced length.

15   Such breakdown products are more likely to engage in non-specific hybridization and are less likely to be effective, relative to their full-length counterparts. Thus, it is desirable to use oligonucleotides that are resistant to degradation in the body and which are able to reach the targeted cells. The present oligonucleotides can be rendered more resistant to degradation *in vivo* by substituting one or more internal

20   artificial internucleotide linkages for the native phosphodiester linkages, for example, by replacing phosphate with sulphur in the linkage. Examples of linkages that may be used include phosphorothioates, methylphosphonates, sulphone, sulphate, ketyl, phosphorodithioates, various phosphoramidates, phosphate esters, bridged phosphorothioates and bridged phosphoramidates. Such

25   examples are illustrative, rather than limiting, since other internucleotide linkages are known in the art. See, for example, Cohen, (1990) *Trends in Biotechnology*. The synthesis of oligonucleotides having one or more of these linkages substituted

61

for the phosphodiester internucleotide linkages is well known in the art, including synthetic pathways for producing oligonucleotides having mixed internucleotide linkages.

5   Oligonucleotides can be made resistant to extension by endogenous enzymes by "capping" or incorporating similar groups on the 5' or 3' terminal nucleotides. A reagent for capping is commercially available as Amino-Link II™ from Applied BioSystems Inc, Foster City, CA. Methods for capping are described, for example, by Shaw *et al* (1991) *Nucleic Acids Res.* **19**, 747-750 and Agrawal *et al*

10   (1991) *Proc. Natl. Acad. Sci. USA* **88(17)**, 7595-7599, the teachings of which are hereby incorporated herein by reference.

A further method of making oligonucleotides resistant to nuclease attack is for them to be "self-stabilised" as described by Tang *et al* (1993) *Nucl. Acids Res.* **21**,

15   2729-2735 incorporated herein by reference. Self-stabilised oligonucleotides have hairpin loop structures at their 3' ends, and show increased resistance to degradation by snake venom phosphodiesterase, DNA polymerase I and fetal bovine serum. The self-stabilised region of the oligonucleotide does not interfere in hybridization with complementary nucleic acids, and pharmacokinetic and

20   stability studies in mice have shown increased *in vivo* persistence of self-stabilised oligonucleotides with respect to their linear counterparts.

It will be appreciated that antisense agents also include larger molecules which bind to said interacting polypeptide mRNA or genes and substantially prevent

25   expression of said interacting polypeptide mRNA or genes and substantially prevent expression of said interacting polypeptide. Thus, expression of an

antisense molecule which is substantially complementary to said interacting polypeptide is envisaged as part of the invention.

The said larger molecules may be expressed from any suitable genetic construct as

5   is described below and delivered to the patient. Typically, the genetic construct which expresses the antisense molecule comprises at least a portion of the said interacting polypeptide coding sequence operatively linked to a promoter which can express the antisense molecule in the cell. Suitable promoters will be known to those skilled in the art, and may include promoters for ubiquitously expressed,

10   for example housekeeping genes or for tissue-specific genes, depending upon where it is desired to express the antisense molecule.

Although the genetic construct can be DNA or RNA it is preferred if it is DNA.

15   Preferably, the genetic construct is adapted for delivery to a human cell.

Means and methods of introducing a genetic construct into a cell in an animal body are known in the art. For example, the constructs of the invention may be introduced into the cells by any convenient method, for example methods

20   involving retroviruses, so that the construct is inserted into the genome of the (dividing) cell.

Other methods involve simple delivery of the construct into the cell for expression therein either for a limited time or, following integration into the

25   genome, for a longer time. An example of the latter approach includes liposomes (Nässander *et al* (1992) *Cancer Res.* **52**, 646-653). Other methods of delivery include adenoviruses carrying external DNA via an antibody-polylysine

63

bridge (see Curiel *Prog. Med. Virol.* **40**, 1-18) and transferrin-polycation conjugates as carriers (Wagner *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3410-3414). The DNA may also be delivered by adenovirus wherein it is present within the adenovirus particle. It will be appreciated that "naked DNA"

5   and DNA complexed with cationic and neutral lipids may also be useful in introducing the DNA of the invention into cells of the patient to be treated. Non-viral approaches to gene therapy are described in Ledley (1995) *Human Gene Therapy* **6**, 1129-1144. Alternative targeted delivery systems are also known such as the modified adenovirus system described in WO 94/10323

10   wherein, typically, the DNA is carried within the adenovirus, or adenovirus-like, particle. Michael *et al* (1995) *Gene Therapy* **2**, 660-668 describes modification of adenovirus to add a cell-selective moiety into a fibre protein. Mutant adenoviruses which replicate selectively in p53-deficient human tumour cells, such as those described in Bischoff *et al* (1996) *Science* **274**, 373-376 are also

15   useful for delivering the genetic construct of the invention to a cell. Thus, it will be appreciated that a further aspect of the invention provides a virus or virus-like particle comprising a genetic construct of the invention. Other suitable viruses or virus-like particles include HSV, AAV, vaccinia and parvovirus.

20   A ribozyme capable of cleaving the interacting polypeptide RNA or DNA. A gene expressing said ribozyme may be administered in substantially the same and using substantially the same vehicles as for the antisense molecules. Ribozymes which may be encoded in the genomes of the viruses or virus-like particles herein disclosed are described in Cech and Herschlag "Site-specific cleavage of

25   single stranded DNA" US 5,180,818; Altman *et al* "Cleavage of targeted RNA by RNAse P" US 5,168,053, Cantin *et al* "Ribozyme cleavage of HIV-1 RNA" US 5,149,796; Cech *et al* "RNA ribozyme restriction endoribonucleases and

64

methods", US 5,116,742; Been *et al* "RNA ribozyme polymerases, dephosphorylases, restriction endonucleases and methods", US 5,093,246; and Been *et al* "RNA ribozyme polymerases, dephosphorylases, restriction endoribonucleases and methods; cleaves single-stranded RNA at specific site by

5   transesterification", US 4,987,071, all incorporated herein by reference.

The genetic constructs of the invention can be prepared using methods well known in the art.

10   A further aspect of the invention provides a method of modulating, for example enhancing or inhibiting, preferably inhibiting, activin or TGFβ signalling in a cell *in vitro* or *in vivo* wherein the cell is exposed to a polypeptide, molecule, nucleic acid, antibody or compound of the invention. It is preferred that the said polypeptide, molecule, nucleic acid, antibody or compound of the invention is

15   able to enter the cell. Methods of optimising delivery to and uptake of such molecules by a cell are known to those skilled in the art and their use is envisaged here. The cell may be a tumour cell, for example a late stage tumour cell.

20   A further aspect of the invention provides the use of a polypeptide, molecule, polynucleotide, compound or antibody of the invention in the manufacture of a medicament for treatment of a patient in need of modulation, preferably inhibition, of activin or TGFβ signalling. A further aspect of the invention provides a method of treatment of a patient in need of modulation, preferably

25   inhibition, of activin or TGFβ signalling wherein an effective amount of a polypeptide, polynucleotide, compound or antibody of the invention is administered to the patient.

65

A further aspect of the invention provides the use of a polypeptide, molecule, polynucleotide, compound or antibody of the invention in the manufacture of a medicament for treatment of a patient with cancer. A further aspect of the

5 invention provides a method of treatment of a patient with cancer wherein an effective amount of a polypeptide, polynucleotide, compound or antibody of the invention is administered to the patient.

For these and for following aspects of the invention it is preferred that the

10 patient is mammalian. It is further preferred that the patient is human.

TGFβ is believed to be involved, for example, in scarring, tissue regeneration and kidney response to diabetes and therefore inhibition of TGFβ signalling *via* the type-I and type-II receptors may be useful in medicine. Activin type-I and

15 type-II receptors may be mediate activins' roles in regulating endocrine cells from the reproductive system, promoters of erythroid differentiation and in inducing axial mesoderm and anterior structures in vertebrates. Inhibins may have effects antagonistic to those of activins. BMP receptors may be involved in similar processes to TGFβ and activins, and particularly in bone growth and

20 maintenance. TGFβs may be expressed in a wider range of tissues than other members of the superfamily, which may have more specialised roles.

TGFβ is also believed to be involved in carcinogenesis (see, for example Lawrence (1996), cited above) and therefore compounds that inhibit TGFβ and

25 related receptor signalling may be useful in the treatment of cancer. Losses of Smad4 may be particularly associated with pancreatic and colon cancers; these cancers may not require TGFβ for progression. Breast cancer tumours are

66

mentioned by Reiss (1997) *Oncol Res* **9**, 447-457 as having high levels of TGFβ associated with them and promoting tumour progression (see also Oft *et al* (1998) TGFβ signalling is necessary for carcinoma cell invasiveness and metastasis *Curr Biol* **8**, 1243-1252).

5

A further aspect of the invention is the use of a polypeptide, molecule, polynucleotide, compound or antibody of the invention in the manufacture of a medicament for treatment of a patient in need of reducing extracellular matrix deposition, encouraging tissue repair and/or regeneration, tissue remodelling or

10   healing of a wound, for example burn, injury or surgery, or reducing scar tissue formation arising from injury to the brain. A further aspect of the invention is a method of treatment of a patient in need of reducing extracellular matrix deposition, encouraging tissue repair and/or regeneration, tissue remodelling or healing of a wound, for example burn, injury or surgery, or reducing scar tissue

15   formation arising from injury to the brain wherein an effective amount of a polypeptide, molecule, polynucleotide, compound or antibody of the invention is administered to the patient.

Extracellular matrix deposition is a term well known to those skilled in the art,

20   and is described for example in Grande (1997) and Lawrence (1996), cited above. Extracellular matrix components include collagens, fibronectin, tenascin, glycosaminoglycans and proteoglycans. Deposition of such components may lead to rapid wound healing but may also lead to scarring, particularly in the brain. TGFβ may inhibit degradation of the extracellular matrix (for example by

25   inhibiting production of proteases and stimulating the production of specific protease inhibitors.

67

It will be appreciated that the medicament may be applied before surgery. It will be appreciated that the injury may be mechanical injury. It is preferred that it is not reperfusion injury.

5   A still further aspect of the invention is the use of a polypeptide, molecule, polynucleotide, compound or antibody of the invention in the manufacture of a medicament for treatment of a patient with or at risk of end-stage organ failure, pathologic extracellular matrix accumulation, a fibrotic condition, disease states associated with immunosuppression (such as different forms of malignancy,

10   chronic degenerative diseases, and AIDS), diabetic nephropathy, tumour growth, kidney damage (for example obstructive neuropathy, IgA nephropathy or non-inflammatory renal disease) or renal fibrosis. A further aspect of the invention provides a method of treating a patient with or at risk of end-stage organ failure, pathologic extracellular matrix accumulation, a fibrotic condition, disease states

15   associated with immunosuppression (such as different forms of malignancy, chronic degenerative diseases, and AIDS), diabetic nephropathy, tumour growth, kidney damage (for example obstructive neuropathy, IgA nephropathy or non-inflammatory renal disease) or renal fibrosis wherein an effective amount of a polypeptide, molecule, polynucleotide, compound or antibody of the invention is

20   administered to the patient.

The patient may alternatively have, or be at risk of, a form of a disorder of bone growth or homeostasis (such as osteoporosis), arthritis or atherosclerosis in which TGFβ or a related protein (for example an activin, inhibin or BMP) has

25   been implicated, in causing or exacerbating the condition. The patient may be suffering from a TGFβ-related condition as reviewed in Roberts & Sport (1993) Physiological actions and clinical applications of transforming growth factor β

68

(TGFβ) *Growth Factors* 8, 1-9, for example hepatic cirrhosis, idiopathic pulmonary fibrosis, scleroderma, glomerulonephritis, certain forms of rheumatoid arthritis, schistosomiasis or proliferative vitreoretinopathy.

5   The polypeptide, molecule, polynucleotide, compound, antibody, composition or medicament of the invention may be administered in any suitable way, usually parenterally, for example intravenously, intraperitoneally or intravesically, in standard sterile, non-pyrogenic formulations of diluents and carriers.   The polypeptide, molecule, polynucleotide, compound, antibody, composition or
10   medicament of the invention of the invention may also be administered topically, which may be of particular benefit for treatment of surface wounds. The polypeptide, molecule, polynucleotide, compound, antibody, composition or medicament of the invention may also be administered in a localised manner, for example by injection.

15

A further aspect of the invention provides a substantially pure complex comprising (1) a Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix2/3 and/or FAST3 polypeptide. It will be appreciated that the interactions between the components of the complex may be
20   non-covalent interactions and that the complex may not be stable at non-physiological pH or salt concentrations. The complex may be stable and/or isolatable under conditions as described in Example 1 in which the complex may be detected by means of immunoprecipitation and/or band shift assays.

25   A further aspect of the invention provides a preparation comprising (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix2/3 and/or FAST3 polypeptide (in the form of a complex or otherwise) when combined with other components *ex vivo*, said other components not being all of

69

the components found in the cell in which said (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix2/3 and/or FAST3 polypeptide (in the form of a complex or otherwise) are naturally found. The preparation may comprise a polypeptide that stabilises the preparation, for 5 example bovine serum albumin or gelatin.

By "substantially pure" we mean that the complex is substantially free of other proteins. Thus, we include any composition that includes at least 30% of the protein content by weight as the said complex or its components, preferably at 10 least 50%, more preferably at least 70%, still more preferably at least 90% and most preferably at least 95% of the protein content is the said complex or its components.

Thus the substantially pure complex may include a contaminant wherein the 15 contaminant comprises less than 70% of the composition by weight, preferably less than 50% of the composition, more preferably less than 30% of the composition, still more preferably less than 10% of the composition and most preferably less than 5% of the composition by weight.

20 The substantially pure said complex may be combined with other components *ex vivo*, said other components not being all of the components found in the cell in which said complex is naturally found.

The invention will now be described in more detail by reference to the following 25 Figures and Examples.

Figure legends

70

Figure 1.   Activin-responsive transcription via the goosecoid DE.

(A) Activin-responsive transcription via the DE is partially dependent on new protein synthe sis. One-cell embryos were injected with REF-globin internal control together with globin reporters driven by the minimal γ-actin promoter (γA),

5   or by multiple copies of the DE or ARE upstream of the mimimal promoter. Animal caps, cut at St 8 were cultured for 6 h ± activin in the absence or presence of 5 μg/ml cycloheximide. Globin transcripts from reporter genes (Test-globin) or the internal control (REF-globin) were detected by RNase protection (Howell and Hill, 1997). Transcriptional activation was calculated as a ratio of the levels of

10   Test-globin to REF-globin. Activin-induced transcription is expressed as fold inductions. In close agreement with the data shown in this experiment, a similar independent experiment measuring the activin inducibility of the DE gave a 22.1-fold induction in the absence of cycloheximide and 6.1-fold in the presence of cycloheximide.

15   (B) An activin-inducible factor (DEBP) binds the goosecoid DE. Whole cell extracts prepared from St 8 or St 11 embryos or St 11 embryos overexpressing activin, were analysed by bandshift assay using the single DE probe. Activin-inducible factor, DEBP is indicated. Competitor oligonucleotides were used at a 50-fold molar excess over probe where indicated. Below, sequences of wild-type

20   DE and mutant oligonucleotides, where only the altered nucleotides are indicated. The paired-like homeodomain binding site comprising 2 inverted TAAT motifs (Wilson *et al* 1993) is denoted by arrows; a third homeodomain binding site at the 3' end is also indicated by an arrow. Thick dotted line, sequence reminiscent of a half-site for the T-box protein, brachyury (AGGTGTGAAATT) (Kispert *et al*

25   1995), and overlapping this (underlined) is an almost perfect binding site for the

ZFH-1 family of zinc finger homeodomain proteins (AGGTGAGCAA) (Funahashi
et al 1993).

(C) Formation of DEBP requires new protein synthesis. Extracts were made from
uninjected St 8 embryos (lane 1), St 10.5 embryos (lanes 2, 3), or St 10.5 embryos

5   overexpressing activin (lanes 4,5) and analysed by bandshift using the DE probe.
Where indicated, embryos had been pre-incubated in cycloheximide before St 8.

Figure 2.   The effector domain of Smad2 interacts with DEBP.

Whole cell extracts prepared from either St 8 embryos (lanes 1-4), St 10.5 embryos

10   (lanes 5-8) or St 10.5 embryos overexpressing activin (lanes 9-12) were analysed
by bandshift using the DE probe. Extracts were mixed with either purified GST
protein (100 ng) (lanes 4, 8, 12) or 2 concentrations (20 ng and 100 ng) of purified
GSTSmad2C (lanes 2,3, 6,7, 10,11) prior to addition of probe. Open arrow, DEBP;
black arrow, GSTSmad2C associated with DEBP.

15

Figure 3.   Homeodomain proteins, Mixer and Milk, but not Mix.1, interact with
Smad2C

(A) Overexpression of Mixer and Milk in Xenopus embryos mimics the activin
induction of DEBP. Whole cell extracts were prepared from St 10.5 embryos or

20   embryos injected at the 1-cell stage with mRNA encoding myc-tagged Mixer,
Milk, Mix.1, or activin, and DE-binding activity was assayed by bandshift. Anti-
myc antibody or purified GSTSmad2C were added where indicated. Open arrow,
DEBP; gray arrow, supershifted complexes.

(B) Interaction of GSTSmad2C with members of the Mix family and Fast-1. In-

25   vitro translated Mixer, Milk, Mix.1 and Fast.1 were assayed by bandshift for their
interaction with purified GSTSmad2C or GST using the appropriate radiolabelled

72

DE or ARE probes. Open arrow, transcription factors complexed with probe; black arrow, ternary complex with GSTSmad2C.

Figure 4.   Characterization of the Smad Interaction Motif (SIM)

5   (A) Schematics of Mix.1, Mixer and Milk, with the conserved homeodomains and a C-terminal acidic domain indicated. Black box; a region conserved in Milk and Mixer, also present in Xenopus Fast-1 and mouse Fast-2 (expanded below where the black line denotes the boundaries of the conserved sequences). Black shading, identical amino acids; gray shading, similar amino acids. The numbers indicate the

10   positions of these amino acids in the full length sequences of the individual proteins.

       (B) C-terminal deletion mutants of Mixer, Milk and Fast.1 (schematized below) were produced *in vitro* and their interaction with GSTSmad2C assayed by bandshift using the DE or ARE probe as appropriate. Complexes of transcription

15   factors and probe are indicated; black arrow, ternary complex with GSTSmad2C. SIM, Smad interaction motif and DNA-binding domains are indicated. Note that Milk gives rise to two complexes, both of which shift with GSTSmad2C, which correspond to a dimer of Milk and a higher order complex.

       (C) Mutation of the prolines in the PP(T/N)K core motif abolishes the interaction

20   with Smad2C. Full length Mixer or a mutant derivative (Mixer PP mut), in which the 2 prolines in the PP(T/N)K containing motif are mutated to alanines, were produced *in vitro* and assayed for interaction with GSTSmad2C by bandshift using the DE probe.

       (D) Interaction of Mixer and Milk and Fast-1 with Smad2C in solution. [^{35}S]-

25   labelled transcription factors as indicated were incubated with Sepharose-bound GST (lanes 2,5,8,11,14) or GSTSmad2C (lanes 3,6,9,12,15) and bound protein was

73

visualized by SDS-PAGE and autoradiography. A fraction of input protein was analysed for comparison (lanes 1,4,7,10,13).

Figure 5.   The Smad interaction motif is sufficient to interact with Smad2.

5   (A) A peptide containing the SIM of Mixer competes specifically for interaction of Mixer with Smad2C. *In vitro*-translated Mixer was incubated with DE probe alone (lane 1) or in the presence of 1 or 10 pmoles of wild type peptide (lanes 2,3) or mutant peptide (lanes 4,5). GSTSmad2C (20 ng) was included in the reactions in lanes 6-14, with the addition of 0.3, 1, 3 or 10 pmoles of wild type peptide (lanes

10   7-10) or mutant peptide (lanes 11-14). Mixer complexed with probe is indicated; black arrow, ternary complex with GSTSmad2C.

(B) A peptide containing the SIM of Mixer specifically disrupts the formation of ARF.

Whole cell extracts made from activin-injected St 8 embryos were analysed by

15   bandshift assay with the ARE probe in the absence (lane 1) or presence of 10, 30, 60, 100, 200 pmoles wild type peptide (lanes 2-6) or mutant peptide (lanes 7-11). The endogenous ARF complex is indicated.

For peptide sequences see Experimental Procedures.

20   Figure 6.   Mixer and Milk interact with activated Smads *in vivo*

(A) Mixer forms a ligand-dependent complex with Smad2 and Smad4 in solution. Extracts were prepared from NIH3T3 cells transfected with myc-Smad2, myc-Smad4 and either Flag-Fast-1, Flag-Mixer, or a Flag-tagged mutant derivative (Mixer PP mut), which had been incubated ± TGF-β1 (2 ng/ml) for 1h. Extracts

25   were assayed either by immunoprecipitation of complexes with anti-Flag antibody followed by Western blotting with anti-Myc antibody (top panel), or Western

74

blotting the whole extract with anti-Flag antibody (middle panel) or with anti-Myc antibody (bottom panel).

(B and C) Fast-1 and Mixer form ligand-dependent complexes on DNA with endogenous Smad2 and Smad4. Extracts were prepared from NIH3T3 cells
5 transfected with Flag-tagged Fast-1, Mixer or Mixer (PP mut), which had been incubated ± TGF-β1 (2 ng/ml) for 1h. Extracts were analysed by bandshift assay on the ARE (B) or DE (C) probe. Anti-flag, anti-Smad2 or anti-Smad4 antibodies were included in the binding reactions where indicated. In (B) ARF and antibody-supershifted ARF are indicated. In (C), Mixer or Mixer (PP mut) bound to probe,
10 the Mixer-Smad complex and antibody-supershifted Mixer-Smad complex are indicated.

Figure 7.   Mixer and Milk mediate TGF-β-dependent transcriptional activation via the DE

15 (A). NIH3T3 cells were transfected with the CAT reporters, and plasmids expressing transcription factors, Smad2 and Smad4 as indicated. Cells were cultured ± TGF-β1 (2 ng/ml) for 8 hr. Cells were harvested and CAT activity measured relative to lacZ activity from the internal control. The data are from a representative experiment, and similar results were obtained in at least three
20 further independent experiments.

(B) Mixer mediates TGF-β dependent transcriptional activation via the DE in the absence of protein synthesis. NIH3T3 cells were transfected with the (DE)4-globin reporter and REF-globin internal control with or without Mixer expression plasmid. Cells were cultured ±TGF-β1 (2 ng/ml) for 4 hr in the absence or
25 presence of 50 μg/ml cycloheximide. Globin transcripts from the reporter genes

75

(test-globin) or the internal control (REF-globin) were detected by RNase protection and quantitated as in Figure 1.

Figure 8.   The temporal and spatial expression patterns of *Mixer* and *Milk* in
5   Xenopus embryos makes them good candidates for mediating transcription of *goosecoid* in response to an endogenous activin-like signal.

(A) Co-expression of *goosecoid* with *Mixer* and *Milk* at early gastrula stages.Xenopus embryos were fixed at St 10.25 and processed for in situ hybridization with probes against *goosecoid* (*Gsc*), *Mixer* or *Milk* either singly (left
10   panels) or sequentially (right panels). Arrowhead, dorsal lip. *Gsc* mRNA is visualized with deep purple stain. *Mixer* and *Milk* mRNA are visualized with a turquoise stain. In the double in situs the overlapping turquoise *Mixer* or *Milk* staining with the purple *Gsc* staining is evident as dark blue staining in dorsal marginal zone (above the dorsal lip). The weak purple background of these
15   embryos is non-specific staining.

(B) Temporal expression patterns of *Mixer*, *Milk* and *goosecoid* in Xenopus embryos.

Time course of expression of *goosecoid* (*Gsc*), *Mixer*, *Milk* and the *FGF receptor* (*FGFR*) assayed by RNase protection. Embryos were sampled at St 8 and
20   subsequent times indicated. In lanes 9-16 the embryos had been pre-incubated with cycloheximide from 30 min before St 8. The *Milk* probe also detects a highly related mRNA, *Milk-related* which is likely to be *Bix3*, which also has a very well conserved PP(T/N)K-containing SIM (see text; Tada *et al* 1998).

(C) A model showing that TGF-β/activ in activated Smads translocate to the
25   nucleus, where they interact with homeodomain transcription factors, Mixer and Milk through the SIM to activate transcription.

76

(D) A model describing the proposed role of the Mixer/Milk-Smad complexes in the formation of mesoderm and endoderm in early Xenopus embryos. The black arrows denote induction of gene expression; the gray arrows denote activation of protein complexes. Milk-related protein/MilkÿSmad complexes are involved in the
5 initiation of transcription of meso-endodermal genes and Mixer/Milk/BixÿSmad complexes are involved in the maintenance of gene expression. For discussion, see text.

Figure 9. Mixer and Milk mediate TGF-β-dependent transcriptional activation via
10 a single DE.

NIH3T3 cells were transfected with a CAT reporter gene driven by a single copy of the goosecoid DE, and plasmids expressing transcription factors Mixer, Mixer (PP mut), Milk and Fast-1 as shown. Cells were cultured ± TGF-β1 (2 ng/ml) for 8 hr. Cells were harvested and CAT activity measured. The data are from a
15 representative experiment, and similar results were obtained in two further independent experiments.

Figure 10. An activin-inducible factor (DEBP) binds the paired-like homeodomain binding site of the *goosecoid* DE.
20 Whole cell extracts prepared from St 8 or St 11 embryos or St 11 embryos overexpressing activin, were analysed by bandshift assay either on the wild type DE probe, or mutant DE probes as indicated. Activin-inducible factor, DEBP is indicated. Below, sequences of wild-type DE and mutant oligonucleotides, where only the altered nucleotides are indicated. The paired-like homeodomain binding
25 site comprising 2 inverted TAAT motifs (Wilson *et al* 1993) is denoted by arrows; a third homeodomain binding site at the 3' end is also indicated by an arrow. Thick

77

dotted line, sequence reminiscent of a half-site for the T-box protein, brachyury (AGGTGTGAAATT) (Kispert *et al* 1995), and overlapping this (underlined) is an almost perfect binding site for the ZFH-1 family of zinc finger homeodomain proteins (AGGTGAGCAA) (Funahashi *et al* 1993).

5

Figure 11. Mapping the Mixer interaction domain in Smad2.

(Left panel) In-vitro translated Mixer was tested in a bandshift assay, using radiolabelled DE as probe, for its ability to interact with different Smad2 effector domain mutants, produced bacterially as GST fusion proteins. Mixer complexed

10   with probe is indicated; black arrow, ternary complex with GSTSmad2C derivatives. (Right panel) Helix 2 of the Smad2 effector domain is required for the interaction with Mixer. The assay was as above. The effector domain of Smad1 does not interact with Mixer. In the mutant GSTSmad2C (H2 swap) Helix 2 from Smad1 replaces Helix 2 of Smad2. This mutant contains only 4 amino acid

15   changes relative to GSTSmad2C (Shi *et al* 1997) and no longer interacts with Mixer.

Figure 12.   Amino acid sequence of human and *Xenopus* Smad2 and Smad3. *Xenopus* Smad3 is a novel Smad polypeptide.

20

Figure 13. A. Alignment of the SIM in proteins of the Mix and Fast families that are known to interact with Smad2. The regions of the proteins are as follows: Mixer, 273-317; ZF Mixer, 228-272; Milk, 309-350; Bix3, 291-332; XFast1, 459-503; HFast1, 316-360; MFast2, 352-396; XFast3, 288-334. In bold are

25   residues that are either completely conserved in all the SIMs or exhibit highly conserved substitutions. Underlined is a pair of amino acids (an acidic residue followed by a hydrophobic residue) that is present in all the SIMs.   Zebrafish

78

Fast-1 has recently been cloned (accession number AF263000) and contains the conserved residues that define the SIM.

B. Alignment of the SIM-containing region of the Mix family members from *Xenopus* and Zebrafish. *Xenopus* and Zebrafish Mixers both contain a SIM; milk and Bix3 also contain a SIM. Bix1, Bix4, Mix.1 and Mix.2 do not contain a SIM. The important conserved residues of the SIM are in bold as in part A.

C. The amino acid sequence of *Xenopus* Fast-3. The forkhead/winged-helix DNA-binding domain is underlined and conserved residues of the SIM are indicated in bold as in part A. The region encompassing the SIM is indicated by a dotted line.

Figure 14. XFast-3 forms an ARF complex in extracts made from *Xenopus* embryos injected with mRNA expressing activin and Flag-tagged XFast-3. Flag-tagged XFast-3 forms a complex in *Xenopus* embryos, ARF2 (Howell et al., 1999) that also contains Smad2 as demonstrated by the observation that the complex supershifts with antibodies against the flag tag on XFast-3 and against Smad2 (lanes 4-6). The complex also contains Smad4 (data not shown). This behaviour is similar to the behaviour of XFast-1, which forms an equivalent complex containing Smad2 and Smad4 (ARF1; lanes 1-3). The high salt extracts containing Flag-tagged XFast-1 were made at 80 min post Stage 8 and those containing Flag-tagged XFast-3 were made 240 min post Stage 8 (Howell et al., 1999).

Figure 15. A mutational analysis of the SIM indicates that the affinity of a Mixer derivative for Smad2 *in vitro* correlates well with the TGF-β inducible transcriptional activity of the Mixer derivative *in vivo*.

A. The sequence of the Mixer SIM indicating the residues that have been mutated to alanine in the single mutations. Underlined is the core of the SIM that is conserved in all functional SIMs (see Figure 13). All mutations have been made in the context of full length Mixer.

5   B. Bandshift analyses to demonstrate the interaction of the Mixer mutants with GSTSmad2C (Example 1 and Germain et al., 2000). The radiolabelled DNA probe is the goosecoid DE. The Mixer derivatives were expressed in reticulocyte lysate. The complex labelled Mixer is a Mixer derivative bound to its binding site in the DE. The black arrow denotes the ternary complex of

10   Mixer/Smad2C/DNA. The titration of GSTSmad2C was in two-fold dilutions. The highest amount added was 20 ng as estimated by Bradford assay. The amounts added were therefore : 20, 10, 5, 2.5, 1.25 and 0.625 ng.

C. TGF-β-induced transcriptional activations of the Mixer derivatives assayed in part B. The cells were NIH3T3s. The reporter was the DE-luciferase reporter

15   which is equivalent to the DE-CAT reporter used previously (Example 1 and Germain et al., 2000) but based on pGL3-basic vector (Promega). TGF-β1 inductions were with 2 ng/ml TGF-β and for 8 h. 50 ng expression plasmid for each mutant derivative and 350 ng reporter plasmid were transfected together with 100 ng EF-lacZ (Example 1 and Germain et al., 2000) as an internal

20   control. Luciferase was quantitated relative to β-Gal activity, and the value for TGF-β induced transcription using wild type Mixer was set at 100. The data are means and standard deviations of 3 independent experiments. The Mixer mutants were all expressed at approximately equal levels in the NIH3T3 cells as determined by bandshift analysis (data not shown).

25

Figure 16.   A. The SIM peptide specifically disrupts the formation of the Smad3/Smad4 complex *in vitro*. The radiolabelled probe was the Smad binding

80

sites from the c-jun promoter (Lehmann et al., 2000). HaCaT cells were treated with 2ng/ml TGF-β1 for 1 h, and nuclear extracts were prepared. 10 μg of nuclear extract was preincubated with antibodies (1μl anti-Smad 3 (Nakao et al 1997) ±10 μg competing peptide); 0.2μg anti-Smad4 (B8, Santa Cruz) or

5   Mixer SIM peptide or mutant Mixer SIM peptide (5, 10, 25, 50, 100 pmoles) for 5 minutes at room temperature prior to probe mix addition. The α-Smad3 competing peptide was pre-incubated with nuclear extract for 5 minutes before antibody addition. The Smad3/4 complex is indicated. The black and white arrows indicate antibody-supershifted complexes.

10   B The SIM peptide specifically disrupts the formation of the XFast-1/Smad2/Smad4 complex *in vitro*, but not the XFast-3/Smad2/Smad4 complex. Whole cell extracts were made from NIH3T3 cells transiently transfected with Flag-XFast-1 and Flag-XFast-3 expression plasmids, that were either untreated or induced with 2ng/ml TGF-β1 for 1 h. Bandshifts were performed using 10 μg

15   extract with the ARE probe (Germain et al., 2000). Extracts were preincubated with antibodies (0.125μg anti-Smad 2/3; Transduction Laboratories or 0.2μg anti-Smad4; B8, Santa Cruz) or Mixer SIM peptide or mutant Mixer SIM peptide (5, 25, 50, 75 pmoles) for 5 minutes at room temperature prior to probe mix addition.

20

Figure 17. A. The SIM peptide specifically disrupts the formation of the Smad3/Smad4 complex *in vivo*. Nuclear extracts were prepared from HaCaT cells treated with 5, 25 or 50 μM Mixer SIM peptide or mutant Mixer SIM peptide for 30 min prior to treatment with 2 ng/ml TGF-β1 for 1 h. The bandshift assay was as in Figure 16A; 10 μg nuclear extract was used in each lane. The antibody supershifts were as in Figure 16B. The black and white arrows indicate antibody-supershifted complexes.

B. The SIM peptide specifically disrupts the formation of the XFast-3/Smad2//Smad4 complex *in vivo*. NIH3T3 cells were transfected with XFast-3. 48 h after transfection cells were treated for 30 min with 50 μM Mixer SIM peptide or mutant Mixer SIM peptide and then treated for 1 h with 2 ng/ml TGF-β1. Nuclear extracts were prepared and 3 μg was used for each lane in the bandshift assay with the ARE probe.

C. The SIM peptide specifically inhibits transcription of the JunB gene *in vivo*. HaCaT cells were treated with 50 μM Mixer SIM peptide or mutant Mixer SIM peptide and then treated for 1 h with 2 ng/ml TGF-β1. Total RNA was extracted and the RNase protection was performed as described (Howell et al., 1999). The γ-actin probe was as described (Enoch et al., 1986). The JunB probe protects amino acids 42-109 of human JunB.

Figure 18. Nucleotide sequence of the XFAST-3 coding region.

Example 1: Homeodomain Transcriptional Partners For Smads

Smads transduce TGF-β signals and participate in transcriptional regulation. We now identify paired-like homeodomain transcription factors of the *Xenopus* Mix family as new partners for activated Smads. We identify a DE-binding protein (DEBP) in *Xenopus* embryos which is synthesized in response to activin and its

82

binding to the paired-like homeodomain site in the DE correlates with activin-induced transcription. DEBP specifically interacts with the effector domain of the activin-activated Smad, Smad2. We demonstrate that two members of the Xenopus Mix family of paired-like homeodomain transcription factors, Mixer (Henry and

5   Melton, 1998) and Milk (Ecochard *et al* 1998) precisely mimic the activity of endogenous DEBP. We demonstrate that Mixer and Milk, but not a third family member, Mix. 1 (Rosa, 1989), directly interact with activin/TGF-β-activated Smad2. This allows recruitment of Smad4 to form an activin/TGFβ-inducible complex that mediates transcriptional activation via the goosecoid DE, the

10   activin-responsive element of the *Xenopus* goosecoid promoter. We have identified a short motif in the C-terminal region of Mixer and Milk, characterized by the sequence PP(T/N)K, which is necessary and sufficient for interaction with the MH2 domain of Smad2. This Smad interaction motif (SIM) is also conserved in the C-terminal regions of the unrelated Smad2-interacting

15   forkhead transcription factors, Fast-1 and Fast-2. Furthermore, we show that Mixer and Milk are expressed in the same cells of the *Xenopus* embryo that express goosecoid, strongly suggesting they are responsible for regulating transcription of goosecoid *in vivo* in response to the endogenous activin-like signals through their interactions with Smads. Our data lead us to propose a

20   model for meso-endoderm formation in *Xenopus* in which these homeodomain transcription factor/Smad complexes play a central role in initiating and maintaining transcription in response to endogenous TGFβ/activin-like signals.

Results

25

83

Activin induced transcription via the distal element of the goosecoid promoter is partly dependent on new protein synthesis

The distal element (DE) in the Xenopus *goosecoid* promoter is a cis-acting element necessary and sufficient to activate transcription in response to activin (Watabe *et*

5   *al* 1995). We first investigated activin-stimulated transcription via the DE in animal cap assays (Howell and Hill, 1997), and compared it with the transcriptional response of the ARE from the Xenopus *Mix.2* promoter, which has a completely different sequence and is known to be controlled by the Fast-1/Smad2/Smad4 complex, ARF (Huang *et al* 1995; Chen *et al* 1996; Chen *et al*

10   1997). We used globin reporter genes with four copies of the DE or three copies of the ARE linked to a minimal promoter and measured transcription by RNase protection assay, quantitating it relative to the activity of a co-injected constitutively active reference globin gene (Howell and Hill, 1997). To get an accurate value for transcription from the TEST-globin plasmid, the amount of

15   transcript from TEST-Globin has to be divided by the amount of transcript from the REF-globin. REF-globin acts as an internal control for injection efficiency, RNA extraction efficiency and as a loading control. The minimal promoter was unresponsive to activin (Figure 1A, left panel). The reporter driven by four DEs responded to activin strongly, and some of this induction was lost in the presence

20   of the protein synthesis inhibitor, cycloheximide (middle panel). The ARE in contrast gave a much higher basal level of transcription, and the activin induction was weaker. As expected, this induction was completely insensitive to cycloheximide (right panel), consistent with it being mediated by the maternal transcription factor complex, ARF.

25

84

From this experiment we conclude that the transcriptional response of the DE to activin has two components: a direct induction mediated by maternal factors, which is insensitive to cycloheximide and a maintenance phase which requires new protein synthesis. A similar behaviour was recently proposed for the related activin-responsive sequence in the zebrafish *goosecoid* promoter (McKendry *et al* 1998).

The *goosecoid* DE binds an activin-inducible factor, DEBP

We used bandshift assays with a radiolabelled single DE oligonucleotide as probe to identify DE-binding factors in the embryo that might be responsible for activin-induced transcription. The DE-binding factor which displayed the expected behaviour is DEBP (DE-binding protein; Figure 1B, lanes 1-3, open arrow). It was absent in extracts prepared from Stage 8 embryos, which are transcriptionally inactive (lane 1). It was present at low levels in extracts from Stage 11 embryos in which endogenous activin-like signaling pathways are operating (lane 2; (Sun *et al* 1999), and highly induced in Stage 11 embryos overexpressing activin (lane 3). Binding of this complex to the DE was specific since it was competed by excess homologous unlabelled probe (lanes 4-6). This complex is probably the same as GAEBP1, shown to bind the related activin-responsive sequence in the zebrafish *goosecoid* promoter (McKendry *et al* 1998).

The DE contains binding sites for several different DNA-binding proteins: a consensus for a paired-like homeodomain protein at its 5' end, consisting of two inverted TAAT motifs separated by 3 nucleotides (Wilson *et al* 1993; McKendry *et al* 1998; arrows, Figure 1B); an additional homeodomain core binding site at the 3' end (arrow); a sequence reminiscent of a half site for the T-box protein,

85

brachyury (dotted line; Kispert *et al* 1995), and overlapping this, a binding site for

the ZFH-1 family of zinc finger homeodomain proteins (underlined; Funahashi *et*

al 1993). We performed competitions with various DE mutants to determine which

of these binding sites was required for DEBP binding.

5   DE m1, which is mutated in the paired-like homeodomain binding site (Watabe *et*

al 1995), competed very poorly for binding (lanes 7-9), indicating that this site is

required. This mutant was also completely inactive in activin-responsive

transcription assays (data not shown; Watabe *et al* 1995; McKendry *et al* 1998),

indicating that this paired-like homeodomain binding site is absolutely required for

10   activin-responsive transcription. DE m2, which is mutated in all three

homeodomain binding sites, did not compete for DEBP binding at all (lanes 10-

12). DE m3, in contrast, which is mutated in the T-box and ZFH-1 and the 3'

homeodomain binding site, competed efficiently for binding (lanes 13-15),

indicating that these sites were not required. The ARE did not compete for DEBP

15   binding (lanes 16-18). Bandshift assays using these mutants as probes were

consistent with these conclusions (data not shown).

Since the activin-responsive transcription via the DE is partly dependent on new

protein synthesis, we asked whether the activin-inducible DEBP also required new

20   protein synthesis for its formation. Indeed, preincubation of the embryos with

cycloheximide prior to initiation of zygotic transcription abolished formation of

DEBP either in Stage 10.5 embryos or Stage 10.5 embryos overexpressing activin

as assayed by bandshift (Figure 1C).

25   Thus the activin-inducible DEBP binds to the paired-like homeodomain binding

site of the DE. The fact that the integrity of this binding site is absolutely required

86

for all the activin-responsive transcription of the DE strongly suggests that DEBP is involved in this (McKendry *et al* 1998). The observation that activin induction of DEBP requires new protein synthesis, indicates that DEBP is most likely to mediate the maintenance phase of transcription of the DE in response to activin.

5   However low levels of maternal DEBP might mediate the component of activin-responsive transcription of the DE that does not require new protein synthesis (McKendry *et al* 1998; see Discussion).

The activin-inducible DEBP can interact with Smad2

10   Activin signals are transduced from activated receptors to the nucleus via a complex of activated Smad2 and Smad4 (Massagué, 1998). We therefore asked whether DEBP might correspond to a Smad/transcription factor complex, by analogy with the Fast-1/Smad complex, ARF. An antibody specific for Smad2 (Nakao *et al* 1997) did not supershift DEBP, indicating that DEBP did not contain

15   endogenous Smad2 and was thus unlikely to be a Smad/transcription factor complex (data not shown). The same Smad2 antibody however could efficiently supershift the ARF complex (data not shown; see Figure 6B).

An alternative possibility was that DEBP was a DNA-binding protein that did

20   interact with activated Smads, but the resulting Smad/DEBP complex was not detectable in our bandshift assays. We therefore investigated whether DEBP could interact with the effector MH2 domain of Smad2 (Smad2C), which is the domain of Smad2 that interacts with Fast-1 in the ARF complex (Chen *et al* 1997; Liu *et al* 1997). Indeed, purified Smad2C, bacterially-expressed as a GST-fusion protein

25   (GSTSmad2C) stoichiometrically supershifted DEBP generated by the endogenous activin-like signals (Figure 2, lanes 5-7) or that generated in response to high

levels of activin signaling (lanes 9-11). This was specific, as GST alone had no

effect (lanes 8, 12). As expected GSTSmad2C alone did not bind the DE probe, as

seen by the lack of binding activity when added to Stage 8 embryo extracts which

do not contain DEBP (lanes 1-3). Thus the supershifts (lanes 6,7,10,11) arise from

5 binding of GSTSmad2C to DEBP, and we conclude that DEBP can interact with

the effector domain of Smad2.

Identification of Smad2-interacting transcription factors that mimic DEBP

The activin-inducible DEBP therefore appears to act as a platform for recruiting

10 Smad2. UV-cross-linking experiments indicated that DEBP corresponded to a

monomer of approximately 45-50 kDa (data not shown). In addition, DEBP binds

the paired-like homeodomain binding site of the DE and is synthesized in response

to activin. A group of transcription factors with precisely these properties are the

paired-like homeodomain proteins of the Mix family. There are seven family

15 members: Mix.1 and the highly related Mix.2 (Rosa, 1989; Vize, 1996), Mixer

(Henry and Melton, 1998), Milk (Ecochard *et al* 1998), also called Bix2 (Tada *et

al* 1998) and three other Bix genes which are highly related to Milk (Tada *et al*

1998). They all have molecular weights of approximately 44 kDa, are first

expressed at the mid to late blastula stage of Xenopus embryogenesis and their

20 expression is known to be induced by activin signaling.

We asked whether overexpression in Xenopus embryos of three different Mix

family members, Mix.1, Mixer and Milk, could mimic the activity of DEBP, both

in DNA-binding specificity and in their ability to interact with Smad2C.

25 Overexpression of myc-tagged Mixer, Milk or Mix.1 alone gave rise to

protein/DNA complexes that co-migrated with the activin-induced DEBP (Figure

88

3A, compare lanes 1, 4, 7, 10 with 13). These protein/DNA complexes could be supershifted with the anti-myc antibody (lanes 5,8,11) indicating the myc-tagged proteins are constituents. Strikingly, only Mixer and Milk have the ability to interact with GSTSmad2C, as shown for endogenous DEBP (compare lanes 6,9, with 3,15). Mix.1 could not associate with GSTSmad2C (lane 12).

We performed an analagous interaction experiment using transcription factors produced *in vitro* by coupled transcription/translation with identical results (Figure 3B, lanes 1-9). As a control for the supershift bandshift assay we also tested the known Smad2-interacting protein, Fast-1 (Chen *et al* 1996), which can be supershifted by GSTSmad2C, but not by GST alone (Figure 3B, lanes 10-12).

Thus Mixer and Milk, but not Mix.1, interact with the effector domain of Smad2, and are therefore good candidates for endogenous DEBP.

Sequences of Smad2 required for interaction with DEBP, Mixer, Milk and Fast-1

We next investigated the sequences in Smad2 required to interact with Mixer, Milk, Fast-1 and endogenous DEBP by assaying a series of Smad2C deletion mutants in the supershift bandshift assay described above (Table 1). Deletion of the phosphorylation sites in the SSMS motif at the extreme C-terminus of Smad2 had no effect on binding to any of the transcription factors (mutant 198-463). Analysis of further N- and C-terminal deletions indicated that the integrity of most of the Smad2 MH2 domain was required for binding to the transcription factors (Table 1). Interestingly Mixer behaved identically to the endogenous DEBP in its interaction with Smad2, whilst Milk behaved like Fast-1 and required additional

89

residues at the C-terminal domain of Smad2C (Table 1, compare mutants 198-445, 198-440, and 198-426). The interaction of the transcription factors with Smad2C was specific, since the equivalent C-terminal region of the BMP-activated Smad, Smad1 (GSTSmad1C) could not interact.

5

The region of Smad2 thought to contact Fast-1 has previously been elucidated and is the α-helix-2 (Chen *et al* 1998). We therefore generated a mutant in which this helix in Smad2 was replaced with the equivalent region of Smad1 (Smad2C H2 Swap; Chen *et al* 1998). This mutant was inactive, indicating that α-helix2 of

10 Smad2 is also required for binding to Mixer, Milk and endogenous DEBP (Table 1).

Identification of a Smad interaction motif

The common property of Smad2 interaction shared by Mixer, Milk and Fast-1

15 prompted us to analyse sequence similarities between these transcription factors. Whereas Mixer and Milk belong to the same family of homeodomain transcription factors, Fast-1 belongs to an unrelated family of winged-helix/forkhead transcription factors (Chen *et al* 1996; Kaufmann and Knöchel, 1996). We identified a short conserved sequence present in the C-terminal region of Mixer,

20 Milk, and Xenopus Fast-1, which was flanked by sequences of no obvious similarity. It is characterized by a completely conserved PP(T/N)K core, flanked by other highly-conserved residues (Figure 4A; black line above sequences). This sequence is also present in human Fast-1 and mouse Fast-2, which also interact with Smad2 (Labbé *et al* 1998; Zhou *et al* 1998; Liu *et al* 1999); Figure 4A).

25 Significantly, the PP(T/N)K core motif is absent in Mix.1, which does not interact with Smad2.

90

To address the potential role of this PP(T/N)K-containing sequence in Smad2 interaction, a series of C-terminal deletion mutants of Mixer, Milk and Fast-1 were produced *in vitro* and assayed by bandshift for their ability to bind the DE and
5   interact with GSTSmad2C. Deletion of the PP(T/N)K-containing sequence in the context of either Mixer, Milk or Fast-1 resulted in the loss of interaction with GSTSmad2C, demonstrating that this sequence is necessary for interaction with Smad2C (Figure 4B). Further C-terminal deletions that impinge on the homeodomains of Mixer or Milk completely abolished DNA binding as expected.
10

The role of the PP(T/N)K core motif for Smad2 interaction was investigated in more detail, by mutating the two conserved prolines of the PP(T/N)K motif to alanine in the context of full length Mixer [Mixer (PP mut)]. This mutation was sufficient to completely abolish the interaction of Mixer with GSTSmad2C,
15   without affecting its DNA binding properties (Figure 4C). This short motif is thus absolutely required for Smad2 interaction.

It was important to establish that these PP(T/N)K-containing transcription factors could also interact with the effector domain of Smad2 in the absence of DNA.
20   Mixer, Milk, and Fast-1 interacted efficiently with Sepharose-bound Smad2C; but Mixer (PP mut) and Mix-1, the family member that does not contain the PP(T/N)K-containing interaction motif, did not (Figure 4D).

Taken together with the results in the previous sections, we have identified Mixer
25   and Milk as Smad2-interacting proteins, and define the PP(T/N)K-containing sequence present in the C-terminal domain of these homeodomain proteins and

91

also present in XFast-1 and mFast-2 as a S̲mad I̲nteraction M̲otif (SIM) essential
for Smad2 interaction.

The Smad interaction motif (SIM) is sufficient to bind Smad2

5   We next investigated whether the SIM was sufficient to interact with Smad2 by
two different assays. First we tested whether a peptide containing 25-amino acids
of Mixer incorporating the SIM (residues 283-307; Figure 4A) could compete with
Mixer for binding Smad2C. Indeed, wild type peptide corresponding to
approximately 10 and 30-fold molar excess over GSTSmad2C was sufficient to

10   inhibit the interaction of Mixer with GSTSmad2C (Figure 5A, lanes 9,10). The
same quantity of the equivalent peptide with the 2 prolines of the PP(T/N)K motif
mutated was ineffective (lanes 13,14). This indicates that the peptide alone is
sufficient to bind Smad2C, thus preventing full length Mixer binding.

15   If, as our data above suggests, the same SIM in Fast-1 is used to recruit active
Smad2 in the complex ARF, then we would expect that the SIM-containing
peptide would be able to disrupt the formation of endogenous ARF. This is exactly
what we observe (Figure 5B). Wild type peptide, but not the mutant, is sufficient to
inhibit the formation of endogenous Xenopus ARF complex (lanes 2-11). Thus the

20   SIM-containing peptide can bind to endogenous Smad2 and inhibit Smad2's
interaction with Fast-1.

Mixer recruits an active Smad complex *in vivo*
We have shown that Mixer and Milk interact with the C-terminal effector domain

25   of Smad2. However, activated Smad2 exists *in vivo* as a complex with Smad4
(Massagué, 1998). We therefore sought direct evidence using co-
immunoprecipitation and bandshift assays that these homeodomain proteins could

92

form stable complexes with ligand-activated Smad2/Smad4 *in vivo*. NIH3T3 cells were used for these experiments since they do not express Mixer or Milk, and this avoided complications of the synthesis of Mix family members in response to activin in *Xenopus* embryo explants. TGF-β was used to stimulate the NIH3T3s as

5   it activates Smad2 and Smad4 in the same way as activin (Liu *et al* 1997) and NIH3T3s respond strongly to TGF-β, and not to activin.

This heterologous system has considerable advantages which allow us to assess the relative importance of Mixer versus the Mixer/Smad complex for transcription via

10   the DE. In particular, 3T3s lack endogenous Mixer/Milk/Bix, but express Smads almost identical to the *Xenopus* Smads, which can be activated in exactly the same way as in a *Xenopus* embryo. This enables us not only to demonstrate that a Mixer/Smad complex forms in response to TGF-β but we can show that this complex is ~ 25-times more transcriptionally active than Mixer alone. Moreover,

15   the double point mutant that doesn't interact with Smads, cannot mediate TGF-β-induced transcription.

These experiments cannot be easily interpreted when performed in animal caps because the results are complicated by the fact that activin induces Mixer/Milk/Bix

20   expression. For instance, the induction of these endogenous genes makes it very difficult to interpret the effect of any mutant derivative, such as Mixer PP mutant.

Figure 6A shows a co-immunoprecipitation assay in which Flag-tagged, Mixer, Mixer (PP mut) or Fast-1 were immunoprecipitated from cells incubated for 1 hour

25   with or without TGF-β and then Western blotted with anti-myc antibody to detect the presence of co-immunoprecipitating myc-tagged Smads. Equal expression of

93

protein was confirmed by Western blotting using anti-Flag or anti-myc antibody of whole cell extracts (Figure 6A, middle and bottom panels). In these conditions of overexpressed Smads, Fast-1 constitutively interacted with both Smad2 and Smad4 (Figure 6A, top panel; but see below). In contrast, in the absence of ligand, Mixer

5   interacted with Smad2 only, but Mixer clearly associated with both Smad2 and Smad4 after TGF-β stimulation (Figure 6A, top panel). Mutation of the two prolines in the SIM in Mixer (Mixer PP mut) completely abolished the formation of this Mixer/Smad complex *in vivo* (Figure 6A, top panel). Thus Mixer can form a ligand-dependent complex with activated Smad2 and Smad4 *in vivo* in the absence

10   of DNA and this requires the integrity of the SIM.

We next determined by bandshift assay on a single DE probe whether Mixer could form a stable TGF-β inducible complex with endogenous Smads on DNA. As a control for the Smad antibodies we demonstrated that they could supershift the

15   Fast-1/Smad2/Smad4 complex, ARF on the ARE probe (Figure 6B). ARF is strongly ligand-inducible in these conditions (lanes 2,7), and clearly contains Fast-1 and endogenous Smad2 and 4 as shown by antibody supershifts (lanes 1-10; Chen *et al* 1996; Chen *et al* 1997). A Mixer/DNA complex is seen in extracts from cells transfected with Flag-tagged Mixer (Figure 6C, lanes 1-8). In addition, a

20   strong TGF-β-induced Mixer-Smad complex was detected with extracts made from cells induced with TGF-β for 1 hour (compare lanes 1 and 5). This Mixer-Smad complex contained endogenous Smad2 and 4 as demonstrated by the antibody supershifts (lanes 6-8). We could additionally prove that TGF-β-inducible Mixer/Smad complex must contain Mixer as well as the Smads, since no

25   such complex was formed in cells expressing Flag-tagged Mixer (PP mut), which does not interact with Smads (lanes 9-16).

94

Mixer and Milk confer TGF-β inducibility on the DE

Having demonstrated that Mixer forms a DNA-binding complex with activated endogenous Smads in response to TGF-β, we investigated whether this complex

5   was transcriptionally active. A DE-driven CAT reporter gene was inactive in NIH3T3 cells and did not respond to TGF-β induction (Figure 7A). Co-transfection of Smad2 and Smad4 had no effect, indicating that the Smads could not activate transcription alone. Mixer displayed very little transcriptional activity in the absence of TGF-β. However, it could confer very strong TGF-β-dependent

10   transcriptional activation on the DE (~ 25-fold induction; Figure 7A). In contrast, the mutant of Mixer that does not bind Smad2 (Mixer PP mut) was completely inactive (Figure 7A). This provided strong evidence that TGF-β induction of transcription via Mixer required recruitment of endogenous Smads. This was corroborated by the observation that overexpression of Smad2 and Smad4

15   potentiated transcription via Mixer in the absence of TGF-β stimulation. Milk also conferred TGF-β inducibility on the DE. However, Mix.1 was inactive, consistent with the fact that it does not interact with Smad2 (Figure 7A). These reporter gene assays were performed with four tandem DE elements. Mixer and Milk were also sufficient to confer TGF-β induced transcription onto a single DE, albeit at a lower

20   level (data not shown). TGF-β induced transcription mediated by the homeodomain proteins were stronger than that elicited by Fast-1 on the ARE (Figure 7A; Liu *et al* 1997), which mirrors what we observe in Xenopus animal cap assays (Figure 1A).

25   Given that the TGF-β activation of transcription mediated by Mixer results from Mixer's interaction with the Smads, we would expect it to be independent of new

protein synthesis. We show that this is indeed the case using globin reporter system where mRNA levels are measured directly by RNase protection. TGF-β-induced transcription via the DE is absolutely dependent on Mixer (Figure 7B, lanes 1,2,5,6) and crucially is not decreased when cycloheximide was added at the

5   same time as TGF-β (lanes 5-8). Thus when Mixer is present TGF-β induced transcription does not require on-going protein synthesis.

Mixer and Milk are expressed appropriately to be endogenous inducers of goosecoid

10   If, as we propose, Mixer and/or Milk are endogenous inducers of *goosecoid* then we would expect them to be expressed in the same domain as *goosecoid*. We investigated the spatial expression patterns of *Mixer*, *Milk* and *goosecoid* by whole mount *in situ* hybridisation in St 10.25 embryos (Figure 8A). In these experiments *Mixer* and *Milk* mRNAs are stained turquoise and *goosecoid* mRNA, deep purple.

15   *Goosecoid* is expressed in the dorsal marginal zone (above the dorsal lip - arrow). *Mixer* and *Milk* are expressed more widely. *Mixer* is expressed throughout the marginal zone (prospective mesoderm) and in vegetal cells (prospective endoderm) and *Milk* is expressed in the dorsal and lateral marginal zone, and in vegetal cells. Mixer was previously thought to be expressed exclusively in the endoderm at stage

20   12 (Henry and Melton, 1998). It is possible that the expression we see in the mesoderm at Stage 10.25 is lost at later stages. In the double *in situs*, the overlapping purple stain of the *goosecoid* signal and the turquoise from the *Mixer* or *Milk* gives a dark blue stain, seen in the dorsal marginal zone (Figure 8A).

25   We also addressed the timing of *Mixer* and *Milk* expression in *Xenopus* embryos. *Mixer* and *Milk* are both expressed before the major upregulation of goosecoid expression (Figure 8B, lanes 1-8). In addition, the *Milk* RNase protection probe

96

also detects a *Milk-related* transcript, likely to be derived from the highly-related *Bix* genes (see below and Experimental procedures; Ecochard *et al* 1998; Tada *et al* 1998).

5   Thus the expression patterns of *Mixer* and *Milk* overlap with *goosecoid* in the dorsal marginal zone at early gastrula stages, and *Mixer* and *Milk* are both expressed before *goosecoid* consistent with them being responsible for induction of *goosecoid*.

10   **The role of Mix family members in meso-endoderm induction**

Finally we addressed the timing of *Mixer* and *Milk* expression relative to the timing of production of the endogenous activin-like signal. In *Xenopus* embryos the major secreted mesoderm-inducing activin-like signal is zygotic and requires the maternal transcription factor VegT for its production (Kimelman and Griffin,

15   1998; Zhang *et al* 1998). Maternal transcription factors, as well as being responsible for producing the activin-like signal that gives rise to the active Smad complexes, might also be responsible for the synthesis of the transcription factors the Smads interact with. Inductions by maternal factors such as VegT should be not be abolished by incubating the embryos in cycloheximide prior to Stage 8.

20

The expression of *Mixer* was virtually all abolished by the cycloheximide treatment, suggesting that it is solely induced by zygotic activators (Figure 8B lanes 1-16). By contrast, *Milk* and the *Milk-related* gene were strongly activated in untreated embryos, and some of this activation remained in cycloheximide-treated

25   embryos (Figure 8B lanes 1-16). This suggests that these genes are weakly induced by maternal activators, and their expression is reinforced by zygotic activators. The

97

temporal expression patterns and sensitivity to cycloheximide of the highly related *Bix* genes, *Bix1, 3* and *4*, were identical to *Milk-related* gene in this assay (data not shown). *Milk-related* is most likely to be *Bix3* from the size of the protected fragment in the RNase protection. Bix3 also contains a well conserved PP(T/N)K-

5   containing SIM.

Thus in the embryo, Milk and Milk-related are likely to be the earliest endogenous Mix family partner for Smads to initiate transcription of meso-endodermal genes. A zygotic signal, probably the endogenous activin-like signal (Ecochard *et al*

10   1998; Tada *et al* 1998) induces the synthesis of additional Milk, Milk-related and also Mixer. In fact this is likely to correspond to the DEBP that we detect at Stage 10.5/11. The complexes this Mixer/Milk/Milk-related form with Smads could be responsible for maintaining the activin-induced transcription of meso-endodermal genes (Figure 8D).

15

Discussion

Mixer and Milk recruit Smads to the *goosecoid* DE to regulate activin/TGF-β responsive transcription

In this example we have investigated the mechanism of activin-responsive

20   transcription via the distal element of the Xenopus *goosecoid* promoter. We have shown that paired-like homeodomain transcription factors of the Mix family, Mixer and Milk, but not Mix.1, mediate activin/TGF-β-induced transcription *via* the DE by interacting specifically with the effector domain of Smad2, thereby recruiting active Smad2/Smad4 complexes to this element (Figure 8C). We

25   demonstrate that the molecular basis for the specificity in the Smad2 interaction is the α -helix 2 of the Smad2 MH2 domain (Shi *et al* 1997). We show that Mixer

98

forms a TGF-β-inducible complex with endogenous Smad2 and Smad4 at the *gooesecoid* DE within 1 hour of ligand stimulation, and we can demonstrate that the Smads are essential for transcriptional regulation mediated by this complex, since the Mixer/Smad complex is approximately 25-fold more transcriptionally active than Mixer alone.

5

Our results also reveal that activated Smads are recruited to different promoter elements by a common mechanism. We have identified a short Smad interaction motif (SIM), characterized by the core sequence PP(T/N)K, in the C-terminal region of Mixer and Milk, which is both necessary and sufficient for these proteins to interact with the effector domain of Smad2. Crucially it is also conserved in the C-terminal regions of the winged-helix/forkhead Smad2-interacting proteins: Xenopus Fast-1, human Fast-1 and mouse Fast-2 (Chen *et al* 1996; Chen *et al* 1997; Labbé *et al* 1998; Zhou *et al* 1998). This indicates that transcription factors of completely different DNA-binding specificity recruit activated Smads to distinct promoter elements *via* the same protein-protein interaction. This finding now explains why activin-responsive elements in the promoters of different *Xenopus* genes share so little sequence similarity (Howell and Hill, 1997).

10

15

20   **Activation of transcription by Smad/transcription factor complexes**

The Smads appear to require other transcription factors to recruit them to DNA because they interact with DNA themselves either very weakly (Smad3 and Smad4) or not at all (Smad2; (Shi *et al* 1998; Hill, 1999). There appears to be a broad range of transcription factor/Smad interactions. At one extreme are the functionally cooperative interactions such as that seen between the *Drosophila*

25

99

homeodomain protein, tinman and MAD and MEDEA (Xu *et al* 1998) where no physical contact between the transcription factor and Smads has been reported. At the other extreme is the direct transcription factor-Smad complexes such as that described here and for Fast-1 and also AP-1 family members (Derynck *et al* 1998). As well as forming transcriptionally active complexes, activated Smads may also be able to release repressors from DNA. Recent work suggests that the homeodomain protein Hoxc-8 functions as a repressor, and interaction with activated Smad1 releases Hoxc-8 from DNA (Shi *et al* 1999).

The interaction of Smads with distinct transcription factors must contribute to cell-type specificity of TGF-β responses, allowing specific genes to be up-regulated only in cells where the essential co-operating transcription factor is also expressed. This is likely to be of particular significance in the patterning of the early *Xenopus* embryo. The same signalling pathway will activate different genes in distinct regions of the embryo depending on the particular Smad-recruiting transcription factors expressed by the cells in that region. In addition, differential affinities of specific transcription factors for Smads, coupled with the presence or absence of Smad binding sites on adjacent DNA could allow distinct genes to be activated by different levels of active Smad complexes. This sort of mechanism might underlie the morphogenetic properties of TGF-β family members, whereby different doses of TGF-β ligands elicit different transcriptional responses (Green and Smith, 1990). Determination of the relative affinities of Mixer, Milk, Bix proteins and Fast-1 for Smad complexes will be important to test these ideas.

Regulation of *goosecoid*

100

Previous studies of Xenopus *goosecoid* regulation indicated that there was an activity in the vegetal hemisphere and marginal zone which could regulate transcription via the DE (Watabe *et al* 1995). This was not sufficient for regulation of the *goosecoid* promoter, which additionally required regulation through the PE

5   by a Wnt-induced transcription factor. The combination of these activities confined *goosecoid* expression to the dorsal marginal zone of the early gastrula (Watabe *et al* 1995; Laurent *et al* 1997). We now propose that the activin-induced DE transcriptional activity corresponds to a SIM-containing member(s) of the Mix family, complexed with activated Smad2/Smad4. Our experiments do not at

10   present allow us to distinguish between Mixer, Milk or other Bix-family members as the endogenous transcription factor responsible. In addition a Fast-1/Smad complex may also be involved in the context of the whole *goosecoid* promoter, since a functional Fast-1 binding site was identified in the mouse *goosecoid* promoter, which is largely conserved downstream of the DE in the *Xenopus*

15   promoter (Labbé *et al* 1998). However, this Fast-1 binding site in the mouse promoter is not sufficient for efficient TGF-β/activin induced transcription, and requires adjacent Smad4 binding sites (Labbé *et al* 1998), which are not conserved in the *Xenopus* promoter (Watabe *et al* 1995). It will be important for the future to investigate possible functional interactions between Fast-1, Mix family members

20   and activated Smads on the Xenopus *goosecoid* promoter.

The role of Mixer and Milk in meso-endodermal induction in Xenopus embryos

Previous work had already implicated Mixer and Milk/Bix in endodermal and

25   mesodermal differentiation, based on experiments in which they were overexpressed in prospective ectoderm (animal caps) (Ecochard *et al* 1998; Henry

101

and Melton, 1998; Tada *et al* 1998). However the underlying mechanism was unknown. Our data suggest that Mixer/Milk/Bix have little inherent transcriptional activity, but require bound Smads activated by an endogenous activin-like signal to increase their transcriptional potential and thus activate meso-endodermal genes.

5   We would therefore predict that the family member Mix.1, which does not interact with Smads, would have a different activity *in vivo*. Indeed, in contrast to Mixer and Milk, Mix.1 does not induce endoderm when overexpressed in animal caps (Ecochard *et al* 1998; Henry and Melton, 1998; Tada *et al* 1998).

10   Our interaction data, together with the expression patterns of these homeodomain proteins allows us to propose a model for meso-endodermal formation in the *Xenopus* embryo (Figure 8D). The major activin-like meso-endoderm-inducing activity that would activate Smad2 and Smad4 is zygotic, and requires the maternal transcription factor, VegT for its production (Kimelman and Griffin, 1998; Zhang

15   *et al* 1998). A good candidate for this ligand is the Vg-1-related protein, derrière (Sun *et al* 1999). Our experiments indicated that Milk and Milk-related, which is probably Bix3, are also induced (weakly) in *Xenopus* embryos by a maternal activator (Figure 8D). This could be VegT itself, since the Bix genes have been shown to be VegT targets (Tada *et al* 1998). Thus low levels of Milk and Milk-

20   related would be available to bind the Smad2/4 complexes activated by the zygotic activin-like ligand to initiate transcription of downstream genes like goosecoid (Figure 8D). In addition, there may be low levels of ubiquitously maternally expressed Milk/Bix genes that would account for the cycloheximide-insensitive activin-induced transcription of the DE seen in the animal caps in Figure 1A. Milk

25   and Milk-related and also Mixer are themselves induced by the zygotic activin-like signaling pathway (Ecochard *et al* 1998; Henry and Melton, 1998; Tada *et al*

102

1998). We propose that these proteins would be involved in maintaining transcription in response to the zygotic activin-like ligand through their formation of transcriptionally active complexes with activated Smads (Figure 8D).

5   In conclusion, our data establish members of the Mix family as transcriptional partners for Smads, responsible for mediating activin/TGF-β responsive transcription in the *Xenopus* embryo via paired-like homeodomain binding sites. It is intriguing that there are a number of highly related genes in the family with apparently identical DNA-binding specificity (Ecochard *et al* 1998; Henry and

10   Melton, 1998; Tada *et al* 1998), and similar expression patterns. Understanding how they each contribute to the patterning of the Xenopus embryo will be an important task for the future.

Experimental procedures

15   **Plasmid constructs**

Mix.1 (Rosa, 1989), Milk (Ecochard *et al* 1998) and Mixer (Henry and Melton, 1998) were isolated by PCR from a Stage 11 *Xenopus* cDNA library and their coding sequences and that of Fast-1 (Chen *et al* 1997) were subcloned into pFTX5 (Howell and Hill, 1997) and EF-Flag. Human Smad4 and Xenopus Smad2 were

20   subcloned in EF-Myc. EF-Flag and EF-myc were derivatives of EF-Plink (Hill *et al* 1995). Prolines 290 and 291 of Mixer were mutated to alanines by PCR. In DE$_4$-CAT and ARE$_3$-CAT four copies of the *goosecoid* DE or three copies of the Mix.2 ARE are upstream of the minimal γ–actin promoter driving CAT. In the globin versions, human β-globin replaced CAT (Howell and Hill, 1997). REF-

25   globin was as described (Howell and Hill, 1997). In GSTSmad2C amino acids 198-467 of XSmad2 and in GSTSmad1C amino acids 172-468 of XSmad1 were

103

subcloned into pGEX-KG (Poon *et al* 1993). 5' and 3' deletions of GSTSmad2C were made using standard methods and named according to the positions of the deletion such that GSTSmad2C(198–245) lacks sequence following codon 245, whilst GSTSmad2C (Δ207-245) lacks sequence between codons 208 and 244.

5   Helix 2 of Smad2 was replaced by Helix 2 of Smad1 in GSTSmad2C using PCR. All constructs were verified by sequencing.

Oligonucleotides

1. CTAGCCATTAATCAGATTAACGGTGAGCAATTAGA (DE-top),

10   2. CCGACTAGTATCTGCTGCCCTAAAATGTGTATTCCATGGAAATG (ARE top),

3. CCGGCTAGCTAGGGAGAGAAGGGCAGACATTTCCATGGAATAC (ARE bot),

4. CTAGCCAGTCAGCAGCTGACCGGTGAGCAATTAGA (DE m1 top),

15   5. CTAGCCAGTCATCAGAGTCACGGTGAGCAAGTCGA (DE m2 top),

6. CTAGCCATTAATCAGATTAACTTGTAGCAAGTAGA (DE m3 top),

GST fusion protein purification, GST "pull-downs" and *in vitro* transcription/ translation

20   Expression of GST-fusion proteins, SDS-PAGE and *in vitro* coupled transcription/translation in reticulocyte lysate (Promega) were performed using standard methods. Mixer, Milk and Fast-1 C-terminal deletion mutants were synthesized *in vitro* using linear templates generated by restriction enzyme digestion. For "pull-down" experiments, [35S]-labelled transcription factors were

25   mixed with GST- or GST-Smad2C-Sepharose beads for 2h at 4°C in 20 mM Tris pH7.5, 20% glycerol, 1 mM EDTA, 5 mM MgCl$_2$, 0.1% NP40 and 220 mM NaCl.

104

The beads were washed three times with five bead volumes of binding buffer, and the protein remaining bound to the beads were analysed by SDS-PAGE followed by autoradiography.

5   **Embryo manipulations, RNase protection assays and *in situ* hybridizations**

The production, maintenance and manipulation of *Xenopus* embryos was previously described (Howell and Hill, 1997). mRNA for microinjection was generated *in vitro* (Howell and Hill, 1997) and injected at the 1-cell stage; 200 pg Activin βA mRNA per embryo, 1.5 ng mRNA encoding myc-tagged Mix.1,

10   Mixer, or Milk. When embryos were treated with cycloheximide, it was added at 20 µg/ml in 0.1X NAM 30 min before St 8. RNA isolation and RNase protection assays were performed as described (Howell and Hill, 1997). The antisense probes were as follows: human *β-globin* (Howell and Hill, 1997); *goosecoid* (Blumberg *et al* 1991); Xenopus *FGF receptor*, protecting amino acids 539-580; *Mixer*, amino

15   acids 173-237; *Milk*, amino acids 143-226. The *Milk* probe also detects a smaller product, whose size and expression characteristics are consistent with it being the protected fragment of the highly Milk-related gene *Bix3* (Tada *et al* 1998). Whole mount *in situ* hybridizations were carried out essentially as described (Harland, 1991), using probes against *goosecoid*, *Mixer* and *Milk* which were identical to

20   those used in the RNase protections, either singly or in combination.

Transfections

NIH3T3 cells were transfected using lipofectamine (Gibco BRL). The following amounts of plasmids were used per 6-cm dish for transcriptional assays, 0.5 µg

25   CAT reporters, 0.2 µg of transcription factors, 0.3 µg of EF-Smad2 and EF-Smad4 and 0.5 µg EF-LacZ as an internal control for transfection efficiency, as indicated

in the Figure legends. For globin transcriptional assays, 1 μg globin reporter, 0.45 μg REF-globin and 0.2 μg Mixer were transfected. For immunoprecipitations, two 6 cm plates were transfected with 0.6 μg of each plasmid. For the bandshift assay, one 6-cm plate was transfected with 1.2 μg transcription factor. The amounts of

5 DNA transfected was kept constant by adding control plasmid EF-plink as appropriate. Following transfection, cells were maintained 18 hr in DMEM containing 10% FBS, before induction by TGF-β1 (2 ng/ml, Calbiochem) for times indicated in the Figure legends.

10 **Transcriptional assays**

After induction, cells were lysed in 200 μl of 20 mM Tris-HCl pH7.5, 150 mM NaCl, 1 mM EDTA and 0.5 % NP40. CAT assays were performed exactly as previously described (Hill *et al* 1993). β-galactosidase assays were performed using CDGP (Calbiochem) as a substrate and quantitated spectrophotometrically.

15 RNA was extracted from NIH3T3 cells for the globin assays as described (Hill *et al* 1994) and the RNase protection assays were as above.

Immunoprecipitation

After induction, cells were lysed in 100 μl buffer containing 20 mM Tris HCl pH

20 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 5 mM NaF, 10 mM β-glycerophosphate, 10% glycerol, 1% Triton and protease inhibitors: 10 μg/ml Leupeptin, E-64, Aprotinin, 20 μg/ml Pepstatin, 0.5 mM Benzamidine and 0.4 mM Pefabloc SC. Flag-tagged transcription factors were immunoprecipitated with anti-Flag M2 affinity gel (Sigma) for 2 hr at 4°C, and washed three times with lysis

25 buffer. Immunoprecipitates were separated by SDS 15% polyacrylamide gel electrophoresis and Western blotted with anti-myc antibody 9E10.

106

Bandshift assays and peptides

Bandshift probes corresponding to the ARE (oligonucleotides 2 and 3) and DE (oligonucleotide 1 and its complement) were labelled with $[\alpha^{32}P]dATP$ and

5      $[\alpha^{32}P]dCTP$ by PCR. Competitions were performed with double-stranded oligonucleotides: DE m1, DE m2, DE m3 or ARE (produced by annealing and filling in oligonucleotides 2 and 3). Whole cell *Xenopus* embryo extracts were prepared by homogenizing embryos in buffer (10 μl per embryo) containing 200 mM KCl, 50 mM Tris-HCl pH 7.4, 10% glycerol, 25 mM β-glycerophosphate, 1

10   mM EGTA, 1 mM EDTA, 2 mM DTT, and protease inhibitors as above. Lysates were cleared by repeated centrifugation. Binding reactions were performed with 30 μg of protein extract incubated with 0.2 ng DE probe in 20 μl of buffer containing 140 mM KCl, 8 mM MgCl$_2$, 12.5 mM β-glycerophosphate, 1 mM EGTA, 1 mM EDTA, 1 mM NaF and 0.5 μg poly(dI-dC), 2 mM DTT, and protease inhibitors for

15   20 min at room temperature. 20 ng of purified GST-fusion proteins were used to study interactions with GSTSmad2C. Extracts for the ARF bandshift in Figure 6B were prepared from activin-injected Stage 8 embryos and the bandshift conditions was as described (Huang *et al* 1995). For *in vitro* translated Mix family members, bandshift conditions were as described (Wilson *et al* 1993). For *in vitro* translated

20   Fast-1, the final buffer concentrations were 8 mM Hepes pH 7.6, 90 mM KCl, 5 mM MgCl$_2$, 4 mM β-glycerophosphate, 40 μM EDTA, 40 μM Spermidine, 2 μg poly(dI-dC), 5% glycerol. Extracts from NIH3T3 cells transfected with Flag-tagged transcription factors were prepared as described (Marais *et al* 1993) and final bandshift conditions were: 14 μg total protein in 10 mM Hepes pH 7.5, 15%

25   glycerol, 210 mM KCl, 5.5 mM MgCl$_2$, 0.2% Triton, 5 mM EGTA, 2.5 mM EDTA, 2 μg poly(dI-dC), 0.25 mM DTT and protease inhibitors with 0.2 ng

107

labelled probe. Specific antibodies (1 µl) anti-Smad2 (Nakao *et al* 1997), anti-
Smad4 (B8; Santa Cruz) or anti-Flag were added to the binding reactions. In all
cases electrophoresis was in 5% polyacrylamide gels/0.5X TBE containing 2.5%
glycerol.

5

The wild type SIM-containing peptide used in Figure 5 was:

Biotin.Aminohexanoicacid-

RQIKIWFQNRRMKWKKLLMDFNNFPPNKTITPDMNVRIPPI. The first 16

amino acids are from the helix 3 of *Antennapedia* which allows internalization of

10   these peptides into live cells (Derossi *et al* 1998); the last 25 amino acids are

codons 283-307 of Mixer. The mutant was the same except that the 2 prolines at

positions 26 and 27 were alanines. Peptides were included in the binding reactions

at the concentrations given in the Figure legend.

15   human recombinant activin A was supplied by NHPP (lot 15365-36(1)).

References

20   Blumberg, B., Wright, C. V., De Robertis, E. M., and Cho, K. W. (1991).
Organizer-specific homeobox genes in Xenopus laevis embryos. Science *253*, 194–
196.

Candia, A. F., Watabe, T., Hawley, S. H., Onichtchouk, D., Zhang, Y., Derynck,
R., Niehrs, C., and Cho, K. W. (1997). Cellular interpretation of multiple TGF-β

25   signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling
mediated by Smads. Development *124*, 4467-4480.

108

Chen, X., Rubock, M. J., and Whitman, M. (1996). A transcriptional partner for MAD proteins in TGF-β signalling. Nature *383*, 691-696.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature *389*, 85-89.

Chen, Y. G., Hata, A., Lo, R. S., Wotton, D., Shi, Y., Pavletich, N., and Massagué, J. (1998). Determinants of specificity in TGF-β signal transduction. Genes Dev. *12*, 2144-2152.

Derossi, D., Chassaing, G., and Prochiantz, A. (1998). Trojan peptides: the penetratin system for intracellular delivery. Trends Cell Biol *8*, 84-87.

Derynck, R., Zhang, Y., and Feng, X. H. (1998). Smads: transcriptional activators of TGF-β responses. Cell *95*, 737-740.

Ecochard, V., Cayrol, C., Rey, S., Foulquier, F., Caillol, D., Lemaire, P., and Duprat, A. M. (1998). A novel Xenopus mix-like gene milk involved in the control of the endomesodermal fates. Development *125*, 2577-2585.

Funahashi, J., Sekido, R., Murai, K., Kamachi, Y., and Kondoh, H. (1993). δ-crystallin enhancer binding protein δEF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis. Development *119*, 433-446.

Green, J. B., and Smith, J. C. (1990). Graded changes in dose of a *Xenopus* activin A homologue elicit stepwise transitions in embryonic cell fate. Nature *347*, 391-394.

Gurdon, J. B., Harger, P., Mitchell, A., and Lemaire, P. (1994). Activin signalling and response to a morphogen gradient. Nature *371*, 487-492.

Harland, R., and Gerhart, J. (1997). Formation and function of Spemann's organizer. Ann. Rev. Cell Dev. Biol. *13*, 611-667.

Harland, R. M. (1991). In situ hybridization: an improved whole-mount method for Xenopus embryos. Methods Cell Biol. *36*, 685-695.

Heasman, J. (1997). Patterning the Xenopus blastula. Development *124*, 4179-4191.

Henry, G. L., Brivanlou, I. H., Kessler, D. S., Hemmati-Brivanlou, A., and Melton, D. A. (1996). TGF-β signals and a pre-pattern in Xenopus laevis endodermal development. Development *122*, 1007-1015.

Henry, G. L., and Melton, D. A. (1998). Mixer, a homeobox gene required for endoderm development. Science *281*, 91-96.

Hill, C. S. (1999). The Smads. Int. J. Biochem. Cell Biol. in press.

Hill, C. S., Marais, R., John, S., Wynne, J., Dalton, S., and Treisman, R. (1993). Functional analysis of a growth factor-responsive transcription factor complex. Cell *73*, 395-406.

Hill, C. S., Wynne, J., and Treisman, R. (1994). Serum-regulated transcription by serum response factor (SRF): a novel role for the DNA-binding domain. EMBO J. *13*, 5421-5432.

Hill, C. S., Wynne, J., and Treisman, R. (1995). The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF. Cell *81*, 1159-1170.

Hogan, B. L. (1996). Bone morphogenetic proteins in development. Curr. Opin. Genet. Dev. *6*, 432-438.

Hogan, B. L. M., Blessing, M., Winnier, G. E., Suzuki, N., and Jones, C. M. (1994). Growth factors in development: the role of TGF-β related polypeptide signalling molecules in embryogenesis. Development, Suppl. 53-60.

Howell, M., and Hill, C. S. (1997). XSmad2 directly activates the activin-inducible, dorsal mesoderm gene XFKH1 in Xenopus embryos. EMBO J. *16*, 7411-7421.

Howell, M., Itoh, F., Pierreux, C. E., Valgeirsdottir, S., Itoh, S., ten Dijke, P., and Hill, C. S. (1999). Xenopus Smad4β is the co-Smad component of

110

developmentally-regulated transcription factor complexes responsible for induction of early mesodermal genes. Dev. Biol. in press.

Huang, H.-C., Murtaugh, L. C., Vize, P. D., and Whitman, M. (1995). Identification of a potential regulator of early transcriptional responses to mesoderm inducers in the frog embryo. EMBO J. *14*, 5965-5973.

Kaufmann, E., and Knöchel, W. (1996). Five years on the wings of forkhead. Mech. Dev. *57*, 3-20.

Kimelman, D., and Griffin, K. J. (1998). Mesoderm induction: a postmodern view. Cell *94*, 419-421.

Kingsley, D. M. (1994). The TGF-β superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes Dev. *8*, 133-146.

Kispert, A., Koschorz, B., and Herrmann, B. G. (1995). The T protein encoded by Brachyury is a tissue-specific transcription factor. EMBO J. *14*, 4763-4772.

Labbé, E., Silvestri, C., Hoodless, P. A., Wrana, J. L., and Attisano, L. (1998). Smad2 and Smad3 positively and negatively regulate TGF β-dependent transcription through the forkhead DNA-binding protein FAST2. Mol. Cell *2*, 109-120.

Laurent, M. N., Blitz, I. L., Hashimoto, C., Rothbacher, U., and Cho, K. W. (1997). The Xenopus homeobox gene twin mediates Wnt induction of goosecoid in establishment of Spemann's organizer. Development *124*, 4905-4916.

Liu, B., Dou, C. L., Prabhu, L., and Lai, E. (1999). FAST-2 is a mammalian winged-helix protein which mediates transforming growth factor β signals. Mol. Cell Biol. *19*, 424-430.

Liu, F., Pouponnot, C., and Massagué, J. (1997). Dual role of the Smad4/DPC4 tumor suppressor in TGFβ-inducible transcriptional complexes. Genes Dev. *11*, 3157-3167.

111

Marais, R., Wynne, J., and Treisman, R. (1993). The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. Cell *73*, 381-393.

Massagué, J. (1998). TGF-β signal transduction. Ann. Rev. Biochem. *67*, 753-791.

5   McKendry, R., Harland, R. M., and Stachel, S. E. (1998). Activin-induced factors maintain goosecoid transcription through a paired homeodomain binding site. Dev. Biol. *204*, 172-186.

Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C. H., Miyazono, K., and ten Dijke, P. (1997). TGF-

10   β receptor-mediated signalling through Smad2, Smad3 and Smad4. EMBO J. *16*, 5353-5362.

Poon, R. Y., Yamashita, K., Adamczewski, J. P., Hunt, T., and Shuttleworth, J. (1993). The cdc2-related protein p40MO15 is the catalytic subunit of a protein kinase that can activate p33cdk2 and p34cdc2. EMBO J. *12*, 3123-3132.

15   Rosa, F. M. (1989). Mix.1, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of Xenopus embryos. Cell *57*, 965-974.

Shi, X., Yang, X., Chen, D., Chang, Z., and Cao, X. (1999). Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling. J. Biol.

20   Chem. *274*, 13711-13717.

Shi, Y., Hata, A., Lo, R. S., Massagué, J., and Pavletich, N. P. (1997). A structural basis for mutational inactivation by the tumour suppressor Smad4. Nature *388*, 87-93.

Shi, Y., Wang, Y. F., Jayaraman, L., Yang, H., Massagué, J., and Pavletich, N. P.

25   (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-β signaling. Cell *94*, 585-594.

112

Sun, B. I., Bush, S. M., Collins-Racie, L. A., LaVallie, E. R., DiBlasio-Smith, E. A., Wolfman, N. M., McCoy, J. M., and Sive, H. L. (1999). derrière: a TGF-β family member required for posterior development in Xenopus. Development *126*, 1467-1482.

5   Tada, M., Casey, E. S., Fairclough, L., and Smith, J. C. (1998). Bix1, a direct target of Xenopus T-box genes, causes formation of ventral mesoderm and endoderm. Development *125*, 3997-4006.

Vize, P. D. (1996). DNA sequences mediating the transcriptional response of the *Mix.2* homeobox gene to mesoderm induction. Dev. Biol. *177*, 226-231.

10   Watabe, T., Kim, S., Candia, A., Rothbacher, U., Hashimoto, C., Inoue, K., and Cho, K. W. (1995). Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between Xenopus and mouse. Genes Dev. *9*, 3038-3050.

Whitman, M. (1998). Smads and early developmental signaling by the TGFβ superfamily. Genes Dev. *12*, 2445-2462.

15   Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. (1993). Cooperative dimerization of paired class homeo domains on DNA. Genes Dev. *7*, 2120-2134.

Xu, X., Yin, Z., Hudson, J. B., Ferguson, E. L., and Frasch, M. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. Genes Dev. *12*, 2354-2370.

Zhang, J., Houston, D. W., King, M. L., Payne, C., Wylie, C., and Heasman, J. (1998). The role of maternal VegT in establishing the primary germ layers in Xenopus embryos. Cell *94*, 515-524.

25   Zhou, S., Zawel, L., Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Characterization of human FAST-1, a TGF β and activin signal transducer. Mol. Cell *2*, 121-127.

113

Table 1. Mapping the transcription factor interaction domain in Smad2

GST-fusions	endogenous DEBP	Mixer	Milk	Fast-1
GST	-	-	-	-
[a]Smad2C (198-467)	+	+	+	+
[b]Smad2C (198-463)	+	+	+	+
Smad2C (198-445)	+	+	-	-
Smad2C (198-440)	+	+	-	-
Smad2C (198-426)	+	+	-	-
Smad2C (198-401)	-	-	-	-
Smad2C (198-373)	-	-	-	-
Smad2C (198-345)	-	-	-	-
Smad2C (198-315)	-	-	-	-
Smad2C (198-276)	-	-	-	-
Smad2C (198-245)	-	-	-	-
Smad2C (Δ 207-245)	+	+	+	+
Smad2C (Δ 207-259)	+	+	+	+
[c]Smad2C (Δ 207-268)	+	+	+	+
[c]Smad2C (Δ 207-321)	-	-	-	-
Smad1C	-	-	-	-
Smad2C(H2 swap)	-	-	-	-

Interactions with purified GST fusion proteins were detected by bandshift assay using radiolabelled DE or ARE probes as appropriate. DEBP was derived from whole cell extracts of St 10.5 embryos overexpressing activin, and Mixer, Milk and Fast-1 were produced *in vitro*.

a The Smad2C protein corresponds to residues 198-467 of Smad2.

b The C-terminal phosphorylation sites (S-465 and S-467) are deleted in this mutant.

c The MH2 domain begins at amino acid W-274.

114

Example 2: Further characterisation of the SIM and its activity as an inhibitor of TGFβ responses

We show that Xenopus Bix1 does not interact with GSTSmad2C, and therefore does not have a functional SIM.

5

PPTK does not appear to form part of a functional SIM in the contect of Bix 1, but there are other differences between the SIM region of Bix1 and the Mixer SIM which may be responsible for the apparent inability of Bix1 to interact with GSTSmad2C. However, it remains possible that in other contexts PPTK would
10   be functional.

Further characterisation of the SIM and characterisation of new SIM-containing family members

Figure 13 shows a line up of functional SIMs, including the new Zebrafish
15   Mixer. It also shows the line up of the SIM region from all the known Mix family members. It is clear that the two Mixers, Milk and Bix3 contain a SIM, but Bix1, Bix4, Mix.1 and Mix.2 do not. Experiments discussed below indicate that those that contain recognizable SIMs bind GSTSmad2C and those that do not, do not bind GSTSmad2C. This confirms that the SIM is responsible for the
20   interaction with GSTSmad2C. Zebrafish Fast-1 also has a SIM (not shown). XFast-3 has a SIM that appears to be functional in this protein. Experiments described below mutagenized the SIM in the context of Mixer and show that affinity for GSTSmad2C *in vitro* correlates well with TGF-β-induced transcriptional activity *in vivo*. The SIM is also demonstrated to be sufficient for
25   TGF-β-induced transcriptional activity *in vivo*.

115

The SIM as an inhibitor of TGF-β responses

The SIM peptides work as inhibitors of Smad2/3-transcription factor interactions *in vitro* and *in vivo*. The formation of transcription factor/Smad complexes that contain Smad2 or Smad3 can be inhibited both *in vitro* and *in vivo*. The TGF-β-

5   induced transcriptional activation of the junB gene can be inhibited. This indicates that interfering with Smad/transcription factor interactions *in vivo* inhibits a biological response to TGF-β. The SIM peptide fused to Antennapedia third helix is transported into cells and reaches the nucleus.

10   *TGF-β-responsive reporter cell lines useful for testing inhibitors of TGF-β signalling.*

TGF-β responsive reporter cell lines can be used to test potential inhibitors of TGF-β signalling, for example peptides or small molecules.

15   **1. Further characterisation of the Smad interaction motif (SIM) and characterisation of new SIM containing family members.**

Figure 13 shows SIMs in different members of the FAST and Mix families from different species.

20

Figure 14 relates to characterisation of *Xenopus* FAST3 (XFAST3) and complexes which comprise XFAST3. XFast-3 can also form complexes with Smad2 and Smad4 in tissue culture cells (see Figure 16B). This is a highly cooperative complex that can not be disrupted by the Mixer SIM peptide *in vitro*.

25   However it is destroyed when the peptide is added to the cells *in vivo* prior to TGF-β stimulation, indicating that the peptide can prevent XFast-3/Smad2/Smad4 complexes forming *in vivo* and that the major Smad2/XFast-3

116

interaction is through the SIM (Figure 17B). This is important for the applications of this peptide as an *in vivo* inhibitor of Smad2 activity and therefore TGF-β/activin inducible transcription. It can clearly prevent active complexes forming *in vivo*. In addition, this data suggests that the XFast-3 SIM may be

5   stronger than the Mixer SIM.

The SIM in XFast-3 is functional. XFast-3 made in reticulocyte lysate bound to the ARE will interact with GSTSmad2C to give a supershift using the methods described in Example 1 and Germain et al., 2000. XFast-3 that is C-terminally

10   truncated so that the SIM is no longer present, does not bind GSTSmad2C efficiently in this assay (data not shown). In addition, a fluorescently-labelled SIM peptide derived from XFast-3 fused to the Antennapedia third helix can prevent the formation of a XFast-1/Smad2/Smad4 complex *in vitro* and formation of a DNA-bound Smad3/Smad4-containing complex in vitro. An

15   equivalent mutant peptide cannot. This indicates that the XFast-3 SIM is capable of specifically interacting with both Smad2 and Smad3.

Characterization of the Mix family members with respect to their interaction with GSTSmad2C.

20

We have further studied six Xenopus Mix family members: Mix.1, Mixer, Bix1, Milk (Bix2), Bix3 and Bix4 and Zebrafish Mixer, with respect to the region of their sequence corresponding to the SIM, and their interaction with GSTSmad2C (see Example 1 and Germain et al., 2000 for methods).

25

In a bandshift assay using the DE as a radiolabelled probe, the family members that interact with GSTSmad2C are Mixer, Milk and Bix3 and Zebrafish Mixer.

117

Bix1, Bix4 and Mix.1 do not interact with GSTSmad2C. This correlates precisely with the presence of a recognizable SIM in Xenopus and Zebrafish Mixer and in Milk and Bix3, but not in Mix1, Bix1 or Bix4 (see Figure 13B). This considerably strengthens our idea that the SIM is responsible for recruiting

5   Smad2 to these proteins *in vitro* and *in vivo*.

Mutagenesis of the SIM to indicate which residues are important for GSTSmad2C binding and for TGF-β-inducible transcription.

10   This is addressed by Figure 15. The affinity of a Mixer derivative for Smad2 *in vitro* correlates well with the TGF-β inducible transcriptional activity of the Mixer derivative *in vivo*. The N residue of the PPNK motif appears to be important for binding to GSTSmad2C. Others residues are also clearly important: F287, F290, P291, P292, K294, T295, I296, M300 and P305. The

15   only residue that we tested that had very little effect when mutated to alanine was D299. However, this residue may be important in the context of the M300 (see Figure 13). TGF-β-induced transcriptional activation via Mixer correlates well with Smad2 binding *in vitro*. All the single and double mutants that do not bind GSTSmad2C *in vitro*, are inactive for TGF-β induced transcription *in vivo*.

20   P292A, M300A and P305A which bind GSTSmad2C very weakly *in vitro*, activate either not at all (M300A) or very weakly in vivo. Those that bind GSTSmad2C significantly *in vitro* (albeit weaker than WT Mixer) have significant activity *in vivo*. The only exception appears to be I296A, which appears to bind GSTSmad2C *in vitro* quite well but is almost completely inactive

25   *in vivo*.

118

The SIM alone is sufficient to confer TGF-β inducibility in vivo

We have made a Gal4 DNA-binding domain (Gal4(1-95)) (Sadowski and Ptashne, 1989) fusion of the SIM (residues 283-307 of Mixer) and a mutant version with the two prolines of the PPNK mutated to alanine. We have assayed

5 these molecules for their ability to confer TGF-β inducible transcription on a luciferase reporter gene derived from pGL3-Enhancer (Promega) driven by 5 Gal4 binding sites. The Gal4(1-95)-SIM can confer approximately 8-fold TGF-β-inducible transcription onto this reporter. The Gal4(1-95)-mutant SIM is completely inactive. The TGF-β inducible transcription mediated by Gal4(1-95)-

10 SIM is competed by overexpression of Mixer or Fast-1, but not by Mixer mutated in the two prolines in the PPNK sequence. These data indicate that *in vivo* the SIM is sufficient to bind the active Smads. Interference with the activity of the SIM is therefore expected to inhibit the activity of the Smads.

15 *The SIM as an inhibitor of TGF-β responses*

The Mixer SIM peptide can disrupt formation of Smad2/3-transcription factor DNA complexes *in vitro*. We have tested three complexes in these assays. First, a TGF-β inducible nuclear complex that contains Smad3 and Smad4, probably in

20 conjunction with an unknown transcription factor that binds the Smad-binding element in the c-jun promoter (Lehmann et al., 2000; Wong et al., 1999). Second, the XFast-1/Smad2/Smad4 complex that binds the Mix.2 ARE (Howell et al., 1999) and third, the XFast-3/Smad2/Smad4 complex that binds the Mix.2 ARE (see Figure 14).

25

The data indicate that the Mixer SIM peptide can efficiently disrupt formation of the Smad3/Smad4-containing complex and the XFast-1/Smad2/Smad4 complex

119

in vitro, but not the XFast-3/Smad2/Smad4 complex, perhaps because the interaction of XFast-3 with Smad2 is stronger than the Mixer SIM peptide interaction with Smad2 (see above). Thus the SIM peptide therefore specifically interacts with Smad2 and Smad3 *in vitro*.

5

The SIM peptide is efficiently taken up by HaCaT and NIH3T3 cells.

The Mixer SIM peptide and the mutant Mixer SIM peptide are taken up by the both NIH3T3 cells and by HaCaT cells when incubated in the normal growth media (10% FCS/DMEM). They are fused to the protein transduction domain,

10   the Antennapedia helix 3, which is why they translocate the plasma membrane. Concentrations that have been tested are between 5 μM and 40 μM. The peptide is found throughout the cells in the cytoplasm and nucleus. The SIM peptide, but not the mutant SIM peptide further accumulates in the nucleus when the cells are treated with TGF-β. The explanation for this may be that the SIM peptide is

15   associated with Smad2 and Smad3 which are cytoplasmic in untreated cells. Some peptide will also be uncomplexed. Upon TGF-β stimulation, the Smad2 and Smad3 translocate to the nucleus, and take the associated peptide with them. The mutant SIM peptide does not show this behaviour probably because it cannot bind the Smads.

20

These peptides are therefore taken up efficiently *in vivo* and are not toxic.

Activity of the SIM peptide as an in vivo inhibitor of TGF-β signalling.

25   Having shown that the peptides are efficiently taken up by NIH3T3 and HaCaT cells we tested whether they can interfere with any *in vivo* TGF-β-induced responses (Figure 17). We can inhibit the formation of the TGF-β induced

120

Smad3/Smad4-containing complex that binds the Smad-binding site of the c-jun promoter, when cells have been incubated with the SIM peptide, but not the mutant SIM peptide. We can inhibit formation of the XFast-3/Smad2/Smad4 complex *in vivo* with the SIM peptide and not with the mutant SIM peptide.

5   These data indicate that the SIM peptide binds Smad2 and Smad3 *in vivo* and inhibits these Smads forming active DNA-binding complexes with different trancription factors in a TGF-β-inducible manner. In addition, we demonstrate that the SIM peptide, but not the mutant peptide, inhibits TGF-β induction of the junB gene by approximately 30%. This is very significant, as it demonstrates

10   that the peptide can inhibit physiological TGF-β responses *in vivo*. Since the mechanism by which TGF-β signalling contributes to diseases such as cancer and fibrosis is through its ability to regulate the transcription of target genes, our ability to inhibit TGF-β induction of transcription of target genes indicates that the SIM peptide (or a small molecule with the same activity) may be an efficient

15   method by which to inhibit TGF-β responses.

Taken together these results tell us: 1. The peptide efficiently gets into cells and into the nucleus. 2. It is not toxic. 3. It specifically binds to Smad2 and Smad3 and can inhibit TGF-β responses.

20

Development of TGF-β-responsive reporter cell lines to test inhibitors of TGF-β signalling.

Generation of TGF-β Inducible Stable Reporter gene Cell Lines.

25   We have shown that the distal element (DE) of the goosecoid promoter is TGF-β inducible and that this inducibility is dependent on the presence of Mixer and active Smad2/Smad4 complexes in transiently transfected NIH3T3 cells

121

(Example 1 and Germain et al., 2000). Similarly we have shown that the activin response element (ARE) confers XFast-1 dependent TGF-β inducibility on a CAT reporter gene in NIH3T3 cell transient transfections (Example 1 and Germain et al., 2000).

5

Stable cell lines may be used to assay potential TGF-β signal transduction pathway inhibitors. NIH3T3 and HaCaT cell lines may be employed for this purpose. The DE and ARE elements may be cloned into, for example, the destabilised enhanced green fluorescent protein promoter cloning vector

10 (pdEGFP-1, Clontech) to generate pDEdEGFP-1 and pAREdEGFP-1. These plasmids carry the neomycin drug resistance gene and hence cells transfected stably with these plasmids can be selected for by growth in media containing G418 (Geneticin, Gibco BRL). The DE and ARE may also be cloned into the secreted alkaline phosphatase reporter gene plasmid pSEAP-2 (Clontech), and

15 into luciferase reporter gene plasmid pGL3-basic (Promega). Stable selection of cell lines carrying these reporter genes may be performed by co-transfecting the plasmid TKNeo (Cruzalegui et al., 1999). Following isolation of stable clones carrying these reporter genes, these clones may be transfected with expression plasmids for Mixer (for DE reporters) and XFast-1 (for ARE reporters). Mixer

20 and XFast-1 may be cloned into the episomal eukaryotic expression vector pCEP4 (Invitrogen). This plasmid contains the hygromycin resistance gene, the Epstein-Barr Virus plasmid origin of replication, the EBNA-1 gene and drives expression of the gene of interest from the CMV immediate early promoter. Stable cell lines which carry both the reporter and the transcription factor may be

25 selected for by growth in media containing both G418 and Hygromycin (Boehringer Mannheim). Similar cell lines may also be generated using the TGF-β-inducible reporter driven by 12 Smad-binding sites from the PAI-1 gene

122

(Dennler et al., 1998). In this case, no transcription factors have to be co-expressed.

Effects of the peptide on reporter gene expression may be analysed by preicubating the cell lines with peptide for 30 minutes prior to TGF-β induction for 8 h or longer. GFP production may be assessed using a confocal microscope and SEAP production can be assayed by sampling the media of the cells and measuring SEAP production using the Great escAPE™ SEAP fluorescent detection kit (Clontech) and a microtitre plate benchtop fluorimeter (Perspective Biosystems). Luciferase can be measured in cell extracts using a luminometer. These cell lines may also be used as the basis of screens for cell soluble small compounds that can interfere with the TGF-β signalling pathway.

Testing in animals may also be useful in identifying or characterising compounds, or in assessing their effects. Because of the high level of conservation of the TGFβ signalling system and components thereof, as discussed above, for example between *Xenopus* and human, it may be appropriate to carry out tests on animals that are not transgenic for components of the TGFβ signalling pathway. However, it may be convenient to use transgenic animals, for example a transgenic animal modified to facilitate detection of modulation of the TGFβ signalling pathway, for example in a manner analogous to the reporter cell lines discussed above.

Relevant methods (not included in Figure legends)

The Antennapedia-SIM peptides
Mixer SIM peptide

123

Biotin.Aminohexanoicacid-

<u>RQIKIWFQNRRMKWKK</u>LLMDFNNFPPNKTITPDMNVRIPPI

Mixer SIM mutant peptide

Biotin.Aminohexanoicacid-

5   <u>RQIKIWFQNRRMKWKK</u>LLMDFNNFAANKTITPDMNVRIPPI

XFast-3 SIM peptide

5-FAM-AMINOHEXANOICACID-

<u>RQIKIWFQNRRMKWKK</u>PEVKNAPKDFPPNKTVFDIPVYTGHPGFLA

XFast-3 mutant SIM peptide

10   5-FAM-AMINOHEXANOICACID-

<u>RQIKIWFQNRRMKWKK</u>PEVKNAPKDFAAAKTVFDIPVYTGHPGFLA

where 5-FAM is 5-carboxyfluorescein (C1359 from Molecular Probes).

The Antennapedia third helix is underlined.

15

The peptides were purified by reverse phase HPLC using an Aquapore ODS 20 micron column (Anachem) in 0.08% trifluoroacetic acid in a gradient of acetonitrile.

20   **Bandshift assays, transfections and transcription assays**

Unless stated, bandshift assays and transfections were as described in Germain et al., 2000. The luciferase reporter assays were performed as described by Jonk et al., 1998, and the β-gal assays were performed using CDGP (Calbiochem) as a substrate and were quantitated spectrophotometrically.

25

124

Treatment of cells in vivo with peptides

Peptides were dissolved in water, and added to the growth media of the cells for the stated times.

5   **Detecting SIM peptides in cells in vivo by immunofluorescence in conjunction with Smad2/3**

Mixer SIM peptide or mutant Mixer SIM peptide was added directly to the growth media of the HaCaT or NIH3T3 cells (10%FCS, DMEM) at concentrations of 5-40μM and was incubated for 30 minutes at 37°C. Following

10   this incubation cells were treated or not with TGF-β1 at a concentraion of 2ng/ml for 1 h at 37°C. Cells were then washed 3 times with ice cold PBS and then fixed in a 5% acetic acid solution in ETOH at -20°C for 30 minutes. Cells were then washed twice with PBS at room temperature and were permeabilised for 10 minutes at room temperature in 3% Tween 20. After two further washes

15   in PBS, cells were incubated in a 1% BSA in PBS for 30 minutes at room temperature. Cells were then incubated with streptavidin texas red (Vector laboratories) at a concentration of 10μg/ml in PBS for 30 minutes at room temperature. Following 3 washes in PBS, cells were then incubated in blocking solution (10% FCS, 0.3% BSA, 0.3% Triton-X100 in PBS) for 30 minutes at

20   room temperature. Cells were then incubated in anti-Smad2/3 monoclonal antibody (Transduction laboratories) at a final concentration of 1μg/ml in blocking solution for 1 h at room temperature. Cells were then washed twice in 0.1% Triton X-100 in PBS and once in PBS and were then incubated with FITC conjugated rabbit anti-mouse immunoglobulins (DAKO) diluted 1 in 200 in

25   blocking solution for 30 minutes at room temperature. Cells were then washed twice in 0.1% Triton X-100 in PBS and once in PBS and were then mounted in

125

Vectashield mounting solution (Vector laboratories). Peptide and Smad staining was visualised using a Axiophot confocal microscope and LSM510 software.

Inhibition of the TGF-β signalling pathway in cancer may be useful.

5   The following is a selection of references showing that tumours secrete TGF-β and that it promotes tumour formation and invasiveness *in vitro* and *in vivo*.

Overexpression of TGF-β in human tumours: Derynck et al., 1987 demonstrates that a variety of human tumours overexpress TGF-β. Gomella et al., 1989

10   concerns enhanced expression of TGF-β in renal cell carcimoma. Steiner and Barrack, 1992 indicates that TGF-β is overproduced in prostate cancer.

TGF-β causes increased tumorigenicity and inhibiting TGF-β signalling by various means inhibits tumour invasiveness and metastasis: Welch et al., 1990

15   indicates that TGF-β stimulates mammary adenocarcinoma cell invasion and metastatic potential. Arteaga et al., 1993a indicates that TGF-β can induce estrogen-independent tumorigenicity of human breast cancer cells in athymic mice. Chang et al., 1993 suggests that increased TGF-β expression inhibits cell proliferation in vitro, yet increases tumorigenicity and tumor growth in Meth A

20   sarcoma cells. Arteaga et al., 1993c indicates that Anti-TGF-β antibodies inhibit breast cancer cell tumorigenicity and increase mouse spleen natural killer cell activity and discussed implications for a possible role of tumor cell/host TGF-β interactions in human breast cancer progression. Arteaga et al., 1993b presents evidence for a positive role of TGF-β in human breast cancer cell tumorigenesis.

25   Huang et al., 1995 suggests that TGF-β1 is an autocrine positive regulator of colon carcinoma U9 cells *in vivo*. Cui et al., 1996 indicates that TGF-β inhibits formation of benign skin tumours, but enhances tumour progression to invasive

126

spindle carcinoma in transgenic mice. Oft et al., 1996 indicates that TGF-β1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells. The invasive phenotypes of the cells is entirely dependent on TGF-β signalling and can be inhibited by neutralizing TGF-β

5   antibodies. Oft et al., 1998 indicates that TGF-β signaling is necessary for carcinoma cell invasiveness and metastasis. Several human carcinoma lines lost invasiveness when treated with neutralizing TGF-β antibodies or soluble receptor variants. Portella et al., 1998 indicates that TGF-β is sufficient to significantly enhance tumorigenicity and the maligant and invasive characteristics of the

10   tumor *in vivo*. These conclusions are drawn from experiments in which the TGF-β signalling pathway was inhibited by overexpresison of a dominant negative TGF-β type II receptor. Yin et al., 1999 indicates that TGF-β signaling blockade inhibits parathyroid hormone related protein secretion by breast cancer cells and bone metastases development. Lehmann et al., 2000 suggests that the

15   ERK MAP kinase pathway synergizes with TGF-β in promoting malignacy. This is reversible by treating cells with neutralizing TGF-β antibodies.

20

References

Akhurst, R.J. and Balmain, A. Genetic events and the role of TGFβ in epithelial
tumour progression. *J. Pathol.*, 187, 82-90. (1999)

25

127

Alexander, J., Rothenberg, M., Henry, G. L. and Stainier, D. Y. casanova plays an early and essential role in endoderm formation in zebrafish *Dev. Biol.* 215, 343-357. (1999)

5   Arteaga, C.L., Carty-Dugger, T., Moses, H.L., Hurd, S.D. and Pietenpol, J.A. Transforming growth factor β 1 can induce estrogen-independent tumorigenicity of human breast cancer cells in athymic mice. *Cell Growth Differ*, 4, 193-201. (1993a)

10   Arteaga, C.L., Dugger, T.C., Winnier, A.R. and Forbes, J.T. Evidence for a positive role of transforming growth factor-beta in human breast cancer cell tumorigenesis. *J Cell Biochem Suppl*, 187-193. (1993b)

     Arteaga, C.L., Hurd, S.D., Winnier, A.R., Johnson, M.D., Fendly, B.M. and
15   Forbes, J.T. Anti-transforming growth factor (TGF)-β antibodies inhibit breast cancer cell tumorigenicity and increase mouse spleen natural killer cell activity. Implications for a possible role of tumor cell/host TGF-β interactions in human breast cancer progression. *J Clin Invest*, 92, 2569-2576. (1993c)

20

     Brummel, T., Abdollah, S., Haerry, T. E., Shimell, M., J.Merriam, J., Raftery, L., Wrana, J. L., O'Connor, M. B. The Drosophila activin receptor baboon signals through dSmad2 and controls cell proliferation but not patterning during larval development. *Genes Dev.* 13, 98-111 (1999).

25

     Chang, H.L., Gillett, N., Figari, I., Lopez, A.R., Palladino, M.A. and Derynck, R. Increased transforming growth factor beta expression inhibits

128

cell proliferation in vitro, yet increases tumorigenicity and tumor growth of
Meth A sarcoma cells. *Cancer Res.*, 53, 4391-4398. (1993)

Chen, X., M.J. Rubock, and M. Whitman. A transcriptional partner for MAD
proteins in TGF-β signalling. *Nature* 383, 691-696. (1996)

Cruzalegui, F.H., Cano, E. and Treisman, R. ERK activation induces
phosphorylation of Elk-1 at multiple S/T-P motifs to high stoichiometry.
Oncogene, 18, 7948-7957. (1999)

Cui, W., Fowlis, D.J., Bryson, S., Duffie, E., Ireland, H., Balmain, A. and
Akhurst, R.J. TGFβ1 inhibits the formation of benign skin tumors, but
enhances progression to invasive spindle carcinomas in transgenic mice.
Cell, 86, 531-542. (1996)

Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S. and Gauthier, J.M.
Direct binding of Smad3 and Smad4 to critical TGF beta-inducible
elements in the promoter of human plasminogen activator inhibitor-type 1
gene. *EMBO J.*, 17, 3091-3100. (1998)

Derynck, R., Goeddel, D.V., Ullrich, A., Gutterman, J.U., Williams, R.D.,
Bringman, T.S. and Berger, W.H. Synthesis of messenger RNAs for
transforming growth factors α and β and the epidermal growth factor
receptor by human tumors. *Cancer Res.*, 47, 707-712. (1987)

129

Dick A., Mayr, T. Bauer, H., Meier A., Hammerschmidt, M. Cloning and characterization of zebrafish Smad2, Smad3 and Smad4. *Gene* 246, 69-80 (2000)

5   Ecochard, V., Cayrol, C., Rey, S., Foulquier, F., Caillol, D., Lemaire, P. and Duprat., A.M. A novel Xenopus mix-like gene milk involved in the control of the endomesodermal fates. *Development* 125, 2577-2585. (1998)

Enoch, T., Zinn, T. and Maniatis, T. Activation of the human β-interferon

10   gene requires an interferon-inducible factor. *Mol. Cell Biol.* 6, 801-810. (1986)

Germain, S., Howell, M., Esslemont, G.M. and Hill, C.S. Homeodomain and winged-helix transcription factors recruit activated Smads to distinct

15   promoter elements via a common Smad interaction motif. *Genes Dev.* 14, 435-451.(2000)

Gomella, L.G., Sargent, E.R., Wade, T.P., Anglard, P., Linehan, W.M. and Kasid, A. Expression of transforming growth factor alpha in normal human

20   adult kidney and enhanced expression of transforming growth factors alpha and beta 1 in renal cell carcinoma. *Cancer Res.*, 49, 6972-6975. (1989)

Henry, G.L. and Melton, D.A. Mixer, a homeobox gene required for endoderm development. Science 281, 91-96. (1998).

25

Howell, M., Itoh, F., Pierreux, C.E., Valgeirsdottir, S., Itoh, S., ten Dijke, P. and Hill, C.S. *Xenopus* Smad4β is the co-Smad component of

130

developmentally regulated transcription factor complexes responsible for induction of early mesodermal genes. *Dev. Biol.* 214, 354-369. (1999)

Huang, F., Newman, E., Theodorescu, D., Kerbel, R.S. and Friedman, E.
5   Transforming growth factor beta 1 (TGF beta 1) is an autocrine positive regulator of colon carcinoma U9 cells in vivo as shown by transfection of a TGF beta 1 antisense expression plasmid. *Cell Growth Differ*, 6, 1635-1642. (1995)

10   Jonk, L. J., S. Itoh, C. H. Heldin, P. ten Dijke, and W. Kruijer. Identification and functional characterization of a Smad binding element (SBE) in the JunB promoter that acts as a transforming growth factor-β, activin, and bone morphogenetic protein-inducible enhancer. *J. Biol. Chem.* 273, 21145-21152.(1998)

15

Labbé, E., C. Silvestri, P.A. Hoodless, J.L. Wrana, and L. Attisano. Smad2 and Smad3 positively and negatively regulate TGF-β-dependent transcription through the forkhead DNA-binding protein FAST2. *Mol. Cell* 2, 109-120. (1998)

20

Lehmann, K., Janda, E., Pierreux, C.E., Rytömaa, M., Schulze, A., McMahon, M., Hill, C.S., Beug, H., and Downward, J. Raf induces TGFβ production while blocking its apoptotic but not invasive responses: a mechanism leading to increased malignancy in epithelial cells. *Genes Dev.*
25   (in press). (2000)

131

Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C.H., Miyazono, K. and ten Dijke, P. TGF-β receptor-mediated signalling through Smad2, Smad3 and Smad4. EMBO J. 16, 5353-5362. (1997)

Oft, M., Heider, K.H. and Beug, H. TGFβ signaling is necessary for carcinoma cell invasiveness and metastasis. *Curr. Biol.*, 8, 1243-1252. (1998)

Oft, M., Peli, J., Rudaz, C., Schwarz, H., Beug, H. and Reichmann, E. TGF-β1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells. *Genes Dev.*, 10, 2462-2477. (1996)

Peale, F.V., Sugden, L. and Bothwell, M. Characterization of CMIX, a chicken homeobox gene related to the Xenopus gene Mix.1. *Mech. Dev.* 75, 179-182 (1998)

Pearce, J.J.H. and Evans, M.J. Mml, a mouse Mix-like gene expressed in the primitive streak. *Mech. Dev.* 87, 189-192 (1999)

Portella, G., Cumming, S.A., Liddell, J., Cui, W., Ireland, H., Akhurst, R.J. and Balmain, A. Transforming growth factor-β is essential for spindle cell conversion of mouse skin carcinoma in vivo: implications for tumor invasion. *Cell Growth Differ.*, 9, 393-404. (1998)

132

Rosa, F.M. Mix.1, a homeobox mRNA inducible by mesoderm inducers, is
    expressed mostly in the presumptive endodermal cells of Xenopus embryos.
    Cell 57: 965-974.
    (1989)

5

Sadowski, I. and Ptashne, M. A vector for expressing GAL4(1-147) fusions in
    mammalian cells. *Nucleic Acids Res*, 17, 7539. (1989)

Steiner, M.S. and Barrack, E.R. Transforming growth factor-beta 1
10   overproduction in prostate cancer: effects on growth in vivo and in vitro.
    Mol Endocrinol, 6, 15-25. (1992)

Tada, M., E.S. Casey, L. Fairclough, and J.C. Smith. Bix1, a direct target of
    Xenopus T-box genes, causes formation of ventral mesoderm and
15   endoderm. *Development* 125, 3997-4006.(1998)

Vize, P.D. DNA sequences mediating the transcriptional response of the *Mix.2*
    homeobox gene to mesoderm induction. *Dev. Biol.* 177, 226-231.(1996)

20   Welch, D.R., Fabra, A. and Nakajima, M. Transforming growth factor beta
    stimulates mammary adenocarcinoma cell invasion and metastatic potential.
    Proc. Natl. Acad. Sci. USA 87, 7678-7682. (1990)

Wong, C., Rougier-Chapman, E.M., Frederick, J.P., Datto, M.B., Liberati,
25   N.T., Li, J.M. and Wang, X.F. Smad3-Smad4 and AP-1 complexes
    synergize in transcriptional activation of the c-Jun promoter by
    transforming growth factor β. *Mol. Cell Biol.*, 19, 1821-1830. (1999)

133

Yin, J.J., Selander, K., Chirgwin, J.M., Dallas, M., Grubbs, B.G., Wieser, R., Massague, J., Mundy, G.R. and Guise, T.A. TGF-beta signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development. *J Clin Invest*, 103, 197-206. (1999)

Zhou, S., L. Zawel, C. Lengauer, K.W. Kinzler, and B. Vogelstein. Characterization of human FAST-1, a TGF-β and activin signal transducer. *Mol. Cell* 2, 121-127. (1998)

Example 3: assay formats for Smad2/SIM interactions

Solid phase

Any method which requires the binding of one of the partners to a solid phase and then measurement of th ebinding of the second partner to it may be used. For example, direct (including SPA) or indirect (including ELISA, displacement) radiochemical, enzymatic or fluorescent methods may be used.

ELISA type assays: where either Smad or SIM is chemically or electrostatically bound to a microwell plate and the binding of the partner molecule detected by enzyme linked antibodie(s) to the partner molecule.

Scintillation proximity assays (SPA): where the binding of Smad to Sim is detected by immobilising one partner on a surface treated with SPA scintillant and radiolabelling the other partner such that a signal occurs if the two partners interact.

134

Displacement assays: where the binding of Smad to SIM is detected by immobilising one partner on a surface and measuring the displacement or binding of labelled (radiochemical or fluorescent) partner.

Homogeneous assays

Methods may be used in which the interaction of Smad and SIM is detected in solution.

Fluorescence resonance energy transfer (FRET): measuring the Smad-SIM interaction by labelling each with a different fluor, the fluorescent wavelength of one giving rise to fluorescence in the other only when the partners are in close proximity. Alternatively, by labelling one partner with a fluor and the other with a quenching dye for the fluor. When the partners are bound together the fluorescence is quenched, to be revealed if the interaction is broken.

Fluorescence correlation microscopy: measuring the diffusion time of either labelled Smad2 or SIM in the presence and absence of the other partner by confocal FCS.

Cell assays

Examples of assays using reporter gene constructs are described above. Assays in which phenotypic characteristics are measured may also be performed.

Consistent with the fact that tumour cells have an intact TGF-β signalling pathway, late stage and metastatic tumours actively overexpress TGF-β1 which

135

acts as a potent tumour promoter. It has direct effects on the tumour cells and also indirect effects through its ability to induce angiogenesis, immunosuppression and alterations in stromal tissue plasticity (Akhurst and Balmain, 1999). The direct tumour promoting effects of TGF-β on the epithelial

5   cells have been well characterized and involve an EMT (epithelial-to-mesenchymal transition) in which the cells loose their polarized phenotype, down-regulate epithelial markers such as E-cadherin and become fibroblastoid in character (Akhurst and Balmain, 1999). They become highly invasive in collagen matrices *in vitro* and efficiently form invasive tumours in mice *in vivo*. In several

10   different mouse and human systems, EMT and the formation of tumours *in vivo* has been shown to be reversed by inhibiting TGF-β signalling either with neutralizing antibodies or by overexpressing dominant negative TGF-β type II receptors (Oft *et al* 1998; Oft *et al* 1996; Portella *et al* 1998). This indicates that the maintenance of the tumour phenotype requires TGF-β signalling.

15

In all the systems studied so far, the tumour promoting effects of TGF-β on epithelial cells are dependent on a synergizing ERK-MAP kinase pathway (Akhurst and Balmain, 1999). It is therefore highly likely that signalling through the ERK-MAP kinase pathway is critical for diverting the TGF-β response from

20·   growth arrest to EMT. The epithelial-to-mesenchymal transition (EMT) occurs in the untransformed dog kidney epithelial cell line, MDCK that inducibly express active Raf-1 (Lehmann *et al* 2000). MDCK cells are available from the American Type Culture Collection of Rockville, MD, USA (ATCC), reference No TCL34. The EMT is dependent on autocrine TGF-β signalling. Normal

25   MDCK cells, which have approximately equal levels of Smad2, Smad3 and Smad4, respond to TGF-β by growth arresting and dying by apoptosis. However after 14 days of expressing high levels of active Raf-1, the cells secrete

136

substantial amounts of active TGF-β and become completely fibroblastoid, are invasive in collagen gels, and no longer growth arrest or apoptose in response to TGF-β. They require autocrine TGF-β signalling to maintain this phenotype.

5   The SIM peptides' or test compounds' ability to reverse EMT in Raf-1-expressing MDCK cells may be measured. Neutralizing TGF-β antibodies can inhibit EMT, as can expression of dominant negative TGF-β type II receptor (Lehmann *et al* 2000; Oft *et al* 1998; Oft *et al* 1996). Inhibition of specific Smad interactions may also reverse EMT in these cells. The ability of SIM

10   peptides or test compounds to reverse invasiveness of metastatic carcinoma cells such as the mouse colon carcinoma cells (CT26) which can be reverted to an epithelial phenotype by TGF-β neutralizing antibodies (Oft *et al* 1998) may also be measured. The results from these experiments may indicate whether a peptide or other test compound has potential as an inhibitor of TGF-β-mediated tumour

15   progression, and may indicate how specific the peptide or compound may be.

References:

Akhurst, R.J. and Balmain, A. (1999) *J. Pathol.*, **187**, 82-90.

Lehmann, K *et al* (2000) The Raf/MAP kinase pathway induces TGFb secretion

20   while blocking the apoptotic but not the invasive responses to the autocrine factor: a mechanism leading to increased malignancy in epithelial cells. *Genes Dev* (2000)

Oft, M *et al* (1998) *Curr. Biol* **8**, 1243-1252.

Oft, M *et al* (1996) *Genes Dev* **10**, 2462-2477.

25   Portella *et al* (1998): *Cell Growth Differ* **9**, 393-404.

<u>CLAIMS</u>

1. A polypeptide (interacting polypeptide) capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises a Smad Interaction Motif (SIM) and is less than 32 amino acids in length.

2. A polypeptide capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises the amino acid sequence PP(T/N)K and is less than 32 amino acids in length.

3. A polypeptide comprising the amino acid sequence PP(T/N)K that is less than 32 amino acids in length.

4. A polypeptide capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises a Smad Interaction Motif (SIM), for example the amino acid sequence PP(T/N)K or three out of four residues thereof, and is not full-length *Xenopus* or human FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer, *Xenopus* Bix3 or Bix2.

5. The polypeptide of claim 1 or 4 wherein the SIM comprises at least 8, 9 or 10 of the specified residues (ie not residues designated by an X) of the amino acid sequence D/E-Hyd-$(X)_n$-P-P-(N/T)-K-(T/S)-(I/V)-$(X)_m$-(D/E)-(M/V/I)-$(X)_k$-P

wherein m= 0 to 7; k= 0 to 8 or 12; n = 0 to 15 or 18.

6. The polypeptide of claim 1, 2, 4 or 5 wherein the Smad polypeptide is Smad2 or Smad3.

138

7. The polypeptide of any one of claims 1 to 6 wherein the polypeptide is a transcription factor or a fragment thereof.

5   8. The polypeptide of any one of claims 4 to 7 wherein the polypeptide is less than 100 amino acids in length.

9. The polypeptide of any of the preceding claims wherein the polypeptide is between 4 and about 30 or 35 amino acids in length.

10

10. The polypeptide of any of the preceding claims wherein an acidic amino acid residue is present at a position from 3 to 10 residues C-terminal of the amino acid sequence PP(T/N)K or amino acid sequence corresponding to the PP(T/N)K motif and/or a proline residue is present at a position from 5 to 20

15   residues C-terminal of the amino acid sequence PP(T/N)K K or amino acid sequence corresponding to the PP(T/N)K motif.

11. The polypeptide of any of the preceding claims comprising the amino acid

sequence   PPNKTITPDMNVRIPPI   or   PPNKTITPDMNTIIPQI   or

20   PPNKSVFDVLTSHPGD   or   PPNKSIYDVWVSHPRD   or

PPNKSIYDVWVSHPRD   or   PPNKTVFDIPVYTGHPG   or

PPNKTITPDMNTIIPQI or PPNKTIGPEMKVVIPPL or PPNKSSKRGNTPPW

or   LLMDFNNFPPNKTITPDMNVRIPPI   or

HSNLMMDFPPNKTITPDMNTIIPQI   or

25   LDNMLRAMPPNKSVFDVLTSHPGD   or

LDSLFQGVPPNKSIYDVWVSHPRD   or

LDALFQGVPPNKSIYDVWVSHPRD   or

139

LKNAPSDFPPNKTVFDIPVYTGHPG or HSNLVMEFPPNKTITPDMNTIIPQI

or   LVEYDNFPPNKTIGPEMKVVIPPL   or

ITSDAYSDSCPPPNKSSKRGNTPPW.

5   12.   A   polypeptide   consisting   of   the   amino   acid   sequence

PPNKTITPDMNVRIPPI   or   PPNKTITPDMNTIIPQI   or

PPNKSVFDVLTSHPGD   or   PPNKSIYDVWVSHPRD   or

PPNKSIYDVWVSHPRD   or   PPNKTVFDIPVYTGHPG   or

PPNKTITPDMNTIIPQI or PPNKTIGPEMKVVIPPL or PPNKSSKRGNTPPW

10   or   LLMDFNNFPPNKTITPDMNVRIPPI   or

HSNLMMDFPPNKTITPDMNTIIPQI   or

LDNMLRAMPPNKSVFDVLTSHPGD   or

LDSLFQGVPPNKSIYDVWVSHPRD   or

LDALFQGVPPNKSIYDVWVSHPRD   or

15 LKNAPSDFPPNKTVFDIPVYTGHPG or HSNLVMEFPPNKTITPDMNTIIPQI

or   LVEYDNFPPNKTIGPEMKVVIPPL   or

ITSDAYSDSCPPPNKSSKRGNTPPW.

13.   The polypeptide of any of the preceding claims comprising the amino acid

20 sequence of residues 283 to 307 of Mixer.

14.   The polypeptide of any of the preceding claims wherein the said polypeptide

is a peptidomimetic compound.

25 15. A molecule comprising a polypeptide as defined in any of Claims 1 to 14 and

a further portion, wherein the said molecule is not full-length *Xenopus* or human

140

FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer or *Xenopus* Bix2.

16.   A molecule according to claim 15 wherein the molecule is

5   Biotin.Aminohexanoicacid-

RQIKIWFQNRRMKWKKLLMDFNNFPPNKTITPDMNVRIPPI

or

5-FAM-AMINOHEXANOICACID-

RQIKIWFQNRRMKWKKPEVKNAPKDFPPNKTVFDIPVYTGHPGFLA

10

17. A nucleic acid encoding or capable of expressing a polypeptide or molecule according to any one of claims 1 to 16.

18.   A nucleic acid complementary to a nucleic acid encoding a polypeptide

15   according to any one of claims 1 to 13.

19.   An antibody capable of reacting with a polypeptide according to any one of claims 1 to 14.

20   20.   A method of identifying a polypeptide that is capable of interacting with a Smad polypeptide, comprising examining the sequence of a polypeptide and determining that the polypeptide comprises a Smad Interaction Motif (SIM), for example the amino acid sequence PP(T/N)K or three out of four residues thereof.

25

21.   The method of claim 20 comprising determining that the polypeptide comprises at least 8, 9 or 10 of the specified residues (ie not residues designated

141

by an X) of the amino acid sequence D/E-Hyd-$(X)_n$-P-P-(N/T)-K-(T/S)-(I/V)-$(X)_m$-(D/E)-(M/V/I)-$(X)_k$-P

wherein m= 0 to 7; k= 0 to 8 or 12; n = 0 to 15 or 18.

5   22. The method of claim 20 or 21 comprising determining that the polypeptide comprises the amino acid sequence PP(T/N)K.

23. The method of claim 20, 21 or 22 further comprising determining that an acid amino acid residue is present at a position from 3 to 10 residues C-terminal

10   of the amino acid sequence PP(T/N)K or amino acid sequence corresponding to the PP(T/N)K motif, and/or a proline residue is present at a position from 5 to 20 residues C-terminal of the amino acid sequence PP(T/N)K or amino acid sequence corresponding to the PP(T/N)K motif.

15   24. A method of identifying a compound capable of disrupting or preventing the interaction between a Smad polypeptide and a target polypeptide that is (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with the said Smad polypeptide, the interaction requiring α-helix2 of the said Smad polypeptide or (3) a polypeptide

20   comprising the amino acid sequence PP(T/N)K, the method comprising measuring the ability of the compound to disrupt or prevent the interaction between the Smad polypeptide and a polypeptide or molecule according to any one of claims 1 to 16.

25   25. A compound identified by or identifiable by the method of claim 24 or claim 47.

142

26. A kit of parts comprising a Smad polypeptide and a polypeptide or molecule according to any one of claims 1 to 16.

27. A method of disrupting or preventing the interaction between a Smad polypeptide and a target polypeptide that is (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with the said Smad polypeptide, the interaction requiring α-helix2 of the said Smad polypeptide, the method comprising exposing the Smad polypeptide to a polypeptide or molecule according to any one of claims 1 to 16 or to an antibody according to claim 19 or to a compound according to claim 25.

28. A method of disrupting or preventing the interaction between a Smad polypeptide and a polypeptide comprising the amino acid sequence PP(T/N)K wherein the Smad polypeptide is exposed to a polypeptide or molecule according to any one of claims 1 to 16 or to an antibody according to claim 19 or to a compound according to claim 25.

29. The method of claim 27 or 28 wherein the Smad polypeptide is Smad2 or Smad3.

30. A compound according to claim 25 or polypeptide or molecule according to any one of claims 1 to 16 or nucleic acid according to claim 17 or 18 or antibody according to claim 19 for use in medicine.

31. A method of modulating activin or TGFβ signalling in a cell *in vitro* wherein the cell is exposed to a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 30.

143

32. A method of modulating activin or TGFβ signalling in a cell *in vivo* wherein the cell is exposed is exposed to a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 30.

5

33. The method of claim 31 or 32 wherein the cell is a late stage tumour cell.

34. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 30 in the manufacture of a medicament for treatment of a
10   patient in need of modulation of activin or TGFβ signalling.

35. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 30 in the manufacture of a medicament for treatment of a patient with cancer.

15

36. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 30 in the manufacture of a medicament for treatment of a patient in need of reducing extracellular matrix deposition, encouraging tissue repair and/or regeneration, tissue remodelling or healing of a wound, injury or
20   surgery, or reducing scar tissue formation arising from injury to the brain.

37. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 30 in the manufacture of a medicament for treatment of a patient with or at risk of end-stage organ failure, pathologic extracellular matrix
25   accumulation, a fibrotic condition, disease states associated with immunosuppression (such as different forms of malignancy, chronic degenerative diseases, and AIDS), diabetic nephropathy, tumour growth, kidney damage (for

144

example obstructive neuropathy, IgA nephropathy or non-inflammatory renal disease) or renal fibrosis.
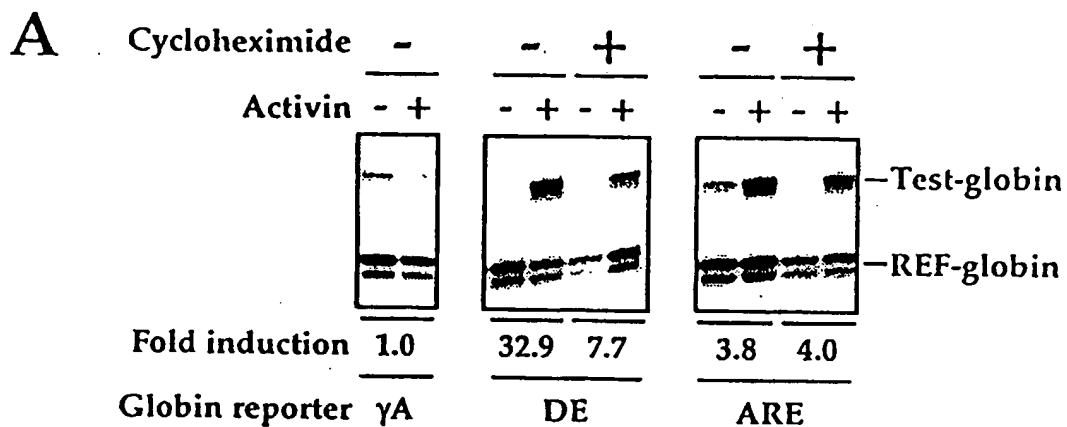
38.   A method of treating a patient in need of modulation of activin or TGFβ
5   signalling, the method comprising administering to the patient an effective amount of a polypeptide, molecule, compound, nucleic acid or antibody as defined in Claim 30.

39.   A method of treating a patient with cancer the method comprising
10   administering to the patient an effective amount of a polypeptide, molecule, compound, nucleic aid or antibody as defined in Claim 30.

40.   A method of reducing extracellular matrix deposition or encouraging tissue repair and/or regeneration, or tissue remodelling or healing of a wound,
15   injury or surgery, or reducing scar tissue formation arising from injury to the brain, the method comprising administering for the patient an effective amount of a polypeptide, molecule, compound, nucleic acid or antibody as defined in Claim 30.

20   41.   A method of treating a disease or condition as defined in Claim 37, the method comprising administering to the patient an effective amount of a polypeptide, molecule, compound, nucleic acid or antibody as defined in Claim 30.

25   42. A substantially pure complex comprising (1) a Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix2/3 and/or FAST3 polypeptide.

145

43. A preparation comprising (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix2/3 and/or FAST3 polypeptide (in the form of a complex or otherwise) when combined with other

5   components *ex vivo*, said other components not being all of the components found in the cell in which said (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix2/3 and/or FAST3 polypeptide (in the form of a complex or otherwise) are naturally found.

10   44. A cell comprising 1) a recombinant polynucleotide suitable for expressing a transcription factor that is capable of interacting with a Smad polypeptide and 2) a recombinant polynucleotide comprising a reporter gene driven by a promoter with a binding site for the said transcription factor.

15   45. A stable cell line cell comprising a reporter gene driven by a promoter with one or more binding sites for an activated Smad, wherein the Smad is activated in the cell by exposure of the cell to TGFβ.

46. The cell according to claim 44 or 45 wherein the reporter gene expresses

20   luciferase, secreted alkaline phosphatase (SEAP), CAT or a green fluorescent protein (GFP).

47. A method of identifying a compound capable of modulating TGFβ-dependent transcription wherein the effect of the compound on expression of the

25   reporter gene in a cell according to claim 44, 45 or 46 is measured, following treatment of the cell with TGFβ.

146

48.   A method of identifying a compound capable of modulating TGFβ-
dependent transcription wherein the effect of the compound on TGFβ-signalling-
dependent invasive behaviour of a stably-transformed cell line cell, for example
in collagen gels, is measured and a compound that reduces invasive behaviour is
5   selected.

49. The method of claim 48 wherein the stably-transformed cell line is a MDCK
cell line that is capable of expressing recombinant active Raf-1.

10

Figure 1

C

Embryo stage	St 8	St 10.5
Cycloheximide	– – +	– +
Activin injected	– – –	+ +

DEBP ⟹

1 2 3 4 5

Figure 1

Figure 2

Embryos	St 8 uninjected	St 10.5 uninjected	St 10.5 + activin
GST	– – – +	– – – +	– – – +
GSTSmad2C	– ◢ –	– ◢ –	– ◢ –

◀ (black arrow)

⟸ DEBP

1 2 3 4   5 6 7 8   9 10 11 12

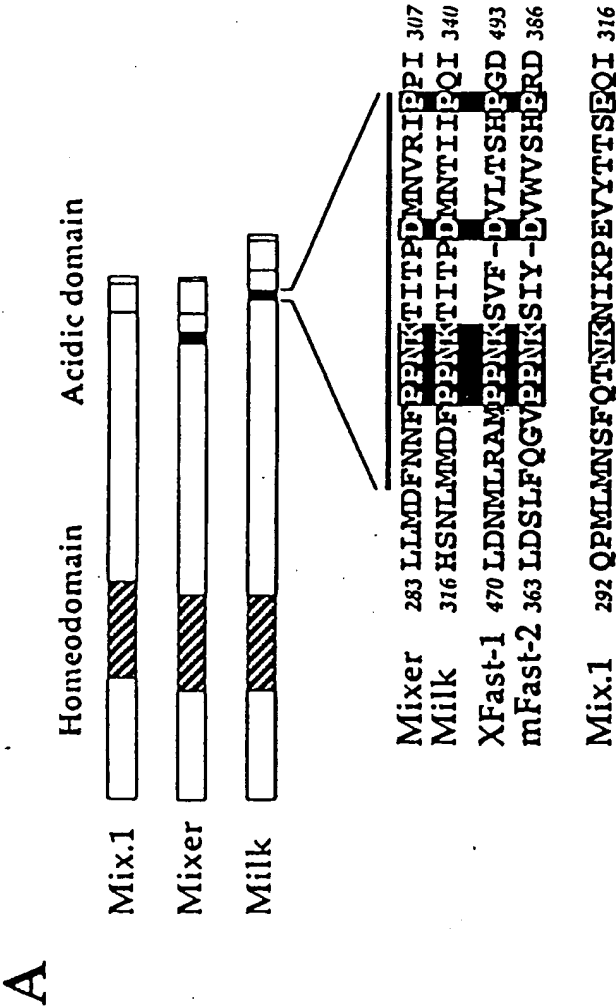2/30

Figure 3

Figure 4

B

Figure 4

5/30

C

GST   - + - - + -
mad2C - - + - - +

Figure 4

Figure 5

Figure 6

A

Smad2+Smad4

Transcription factor

Reporter

B

Transcription factor

Cycloheximide

TGF-β

—Test-globin

—REF-globin

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

Fold induction   0.9   1.0   6.5   10.9

Reporter   (DE)₄-Globin

Figure 7

9/30
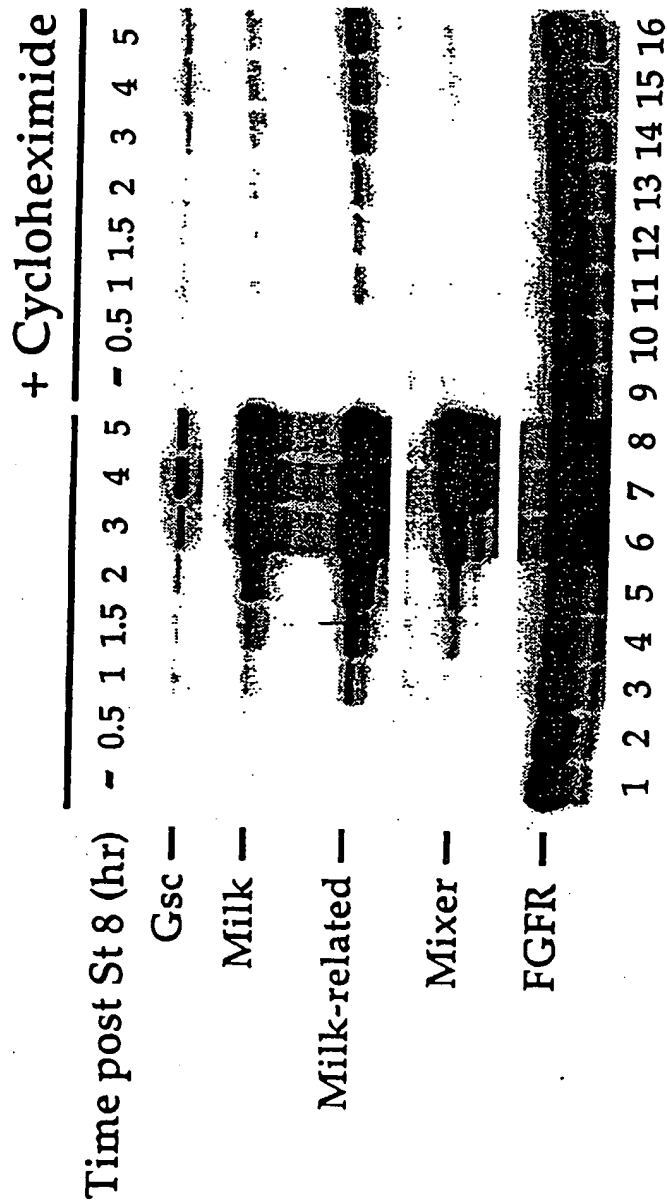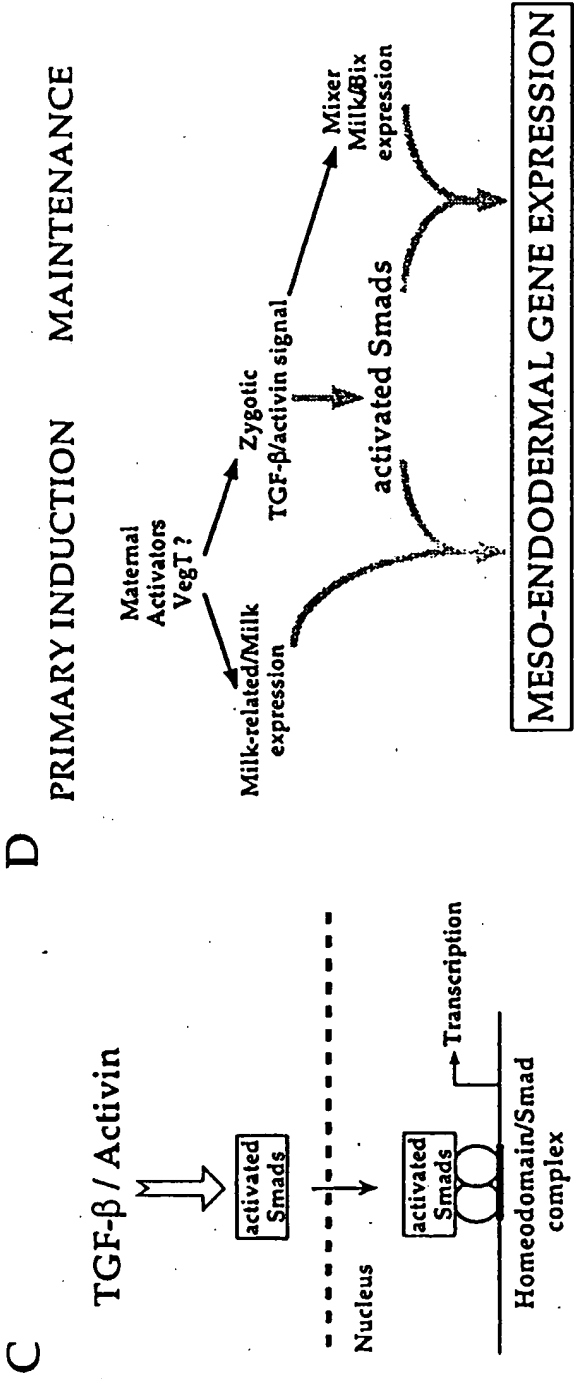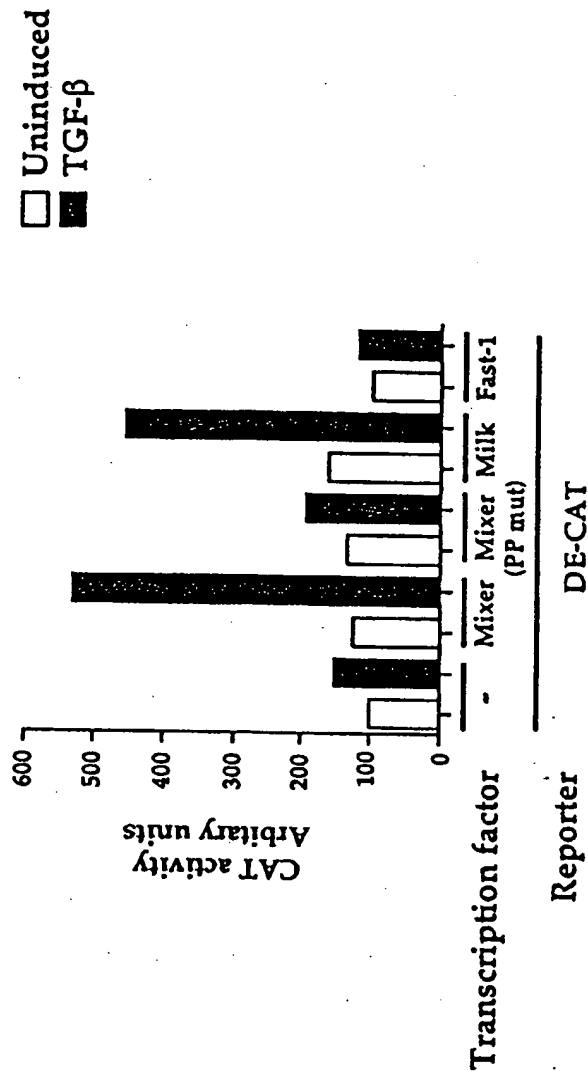
Figure 8

Figure 8

Figure 8

Figure 9

Figure 10

14/30

Figure 11

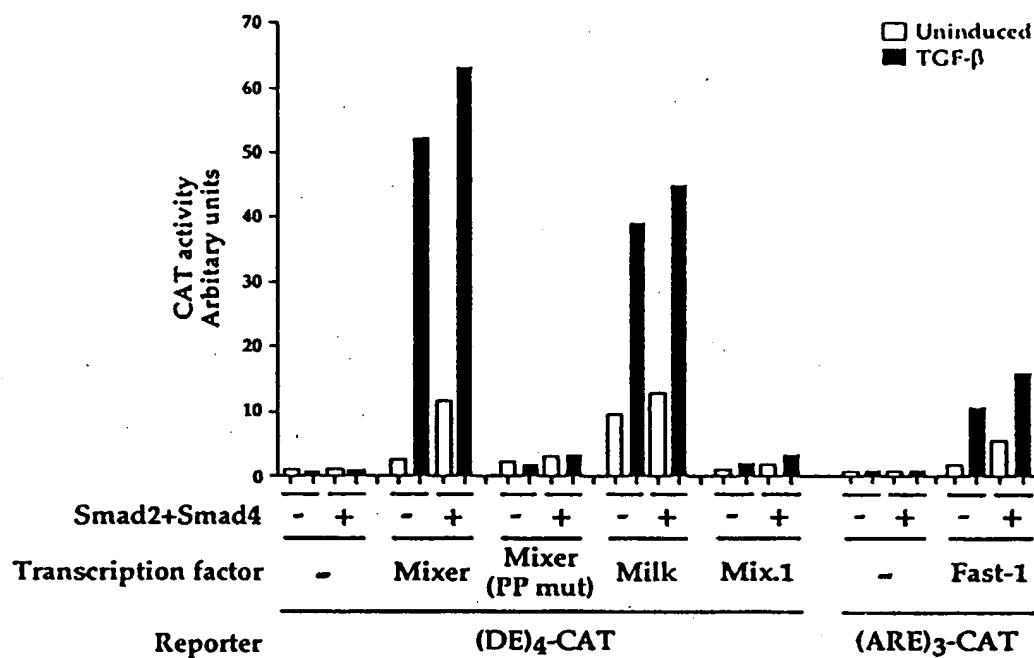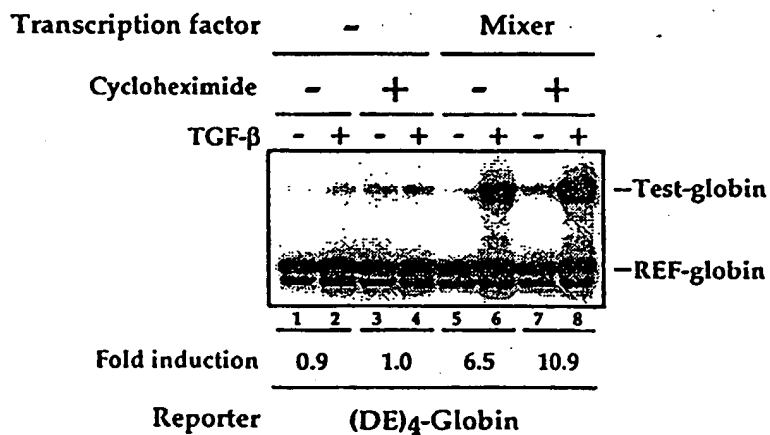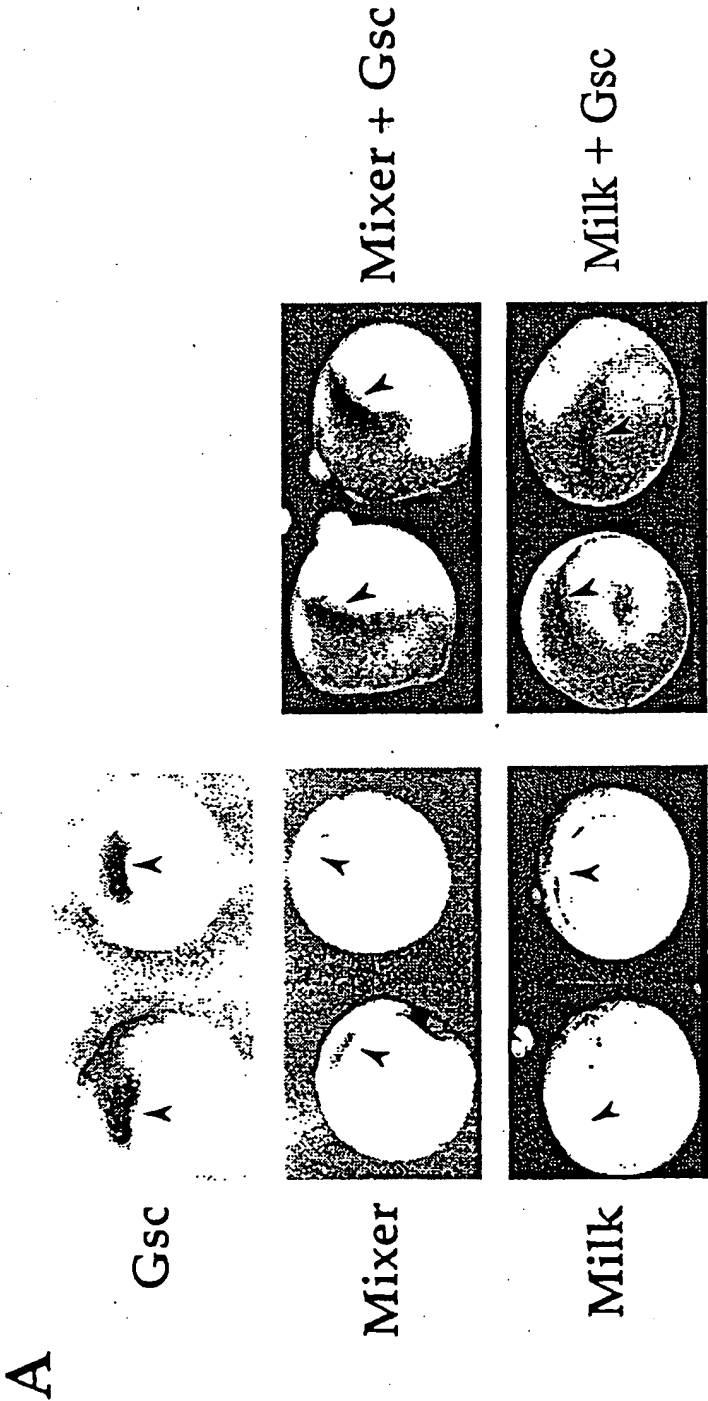XSmad2
MAGSMSSILPFTPPVVKRLLGWKKSASGTTGAGGDEQNGQEEKWCEKAVK
SLVKKLKKTGQLDELEKAITTQNCNTKCVTIPSTCSEIWGLSTANTIDQW
DTTGLYSFSEQTRSLDGRLQVSHRKGLPHVIYCRLWRWPDLHSHHELKAI
ENCEYAFNLKKDEVCVNPYHYQRVETPVLPPVLVPRHTEILTELPPLDDY
THSIPENTNFPAGIEPQSNYIPETPPPGYISEDGETSDQQLNQSMDTGSP
AELSPSTLSPVNHNLDLQPVTYSEPAFWCSIAYYELNQRVGETFHASQPS
LTVDGFTDPSNSERFCLGLLSNVNRNATVEMTRRHIGRGVRLYYIGGEVF
AECLSDSAIFVQSPNCNQRYGWHPATVCKIPPGCNLKIFNNQEFAALLAQ
SVNQGFEAVYQLTRMCTIRMSFVKGWGAEYRRQTVTSTPCWIELHLNGPL
QWLDKVLTQMGSPSVRCSSMS

Hsmad2
MSSILPFTPPVVKRLLGWKKSAGGSGGAGGGEQNGQEEKWCEKAVKSLVK
KLKKTGRLDELEKAITTQNCNTKCVTIPSTCSEIWGLSTPNTIDQWDTTG
LYSFSEQTRSLDGRLQVSHRKGLPHVIYCRLWRWPDLHSHHELKAIENCE
YAFNLKKDEVCVNPYHYQRVETPVLPPVLVPRHTEILTELPPLDDYTHSI
PENTNFPAGIEPQSNYIPETPPPGYISEDGETSDQQLNQSMDTGSPAELS
PTTLSPVNHSLDLQPVTYSEPAFWCSIAYYELNQRVGETFHASQPSLTVD
GFTDPSNSERFCLGLLSNVNRNATVEMTRRHIGRGVRLYYIGGEVFAECL
SDSAIFVQSPNCNQRYGWHPATVCKIPPGCNLKIFNNQEFAALLAQSVNQ
GFEAVYQLTRMCTIRMSFVKGWGAEYRRQTVTSTPCWIELHLNGPLQWLD
KVLTQMGSPSVRCSSMS

Figure 12

Xsmad3
MSSILPFTPPIVKRLLGWKKGEQNGQEEKWCEKAVKSLVKKLKKSGQLDE
LEKALTTQSISTKCITIPRSLDGRLQVSHRKGLPHVIYCRLWRWPDLHSH
HELRAMEVCEYAFSMKKDEVCVNPYHYQRVETPVLPPVLVPRNTEIPAEF
PSLDDYSHSIPENTNFPAGIEPQINYIPETPPPGYLSEDGETSDQMNHSI
DTGSPNLSPNSMSPAHSNMDLQPVTYCEPAFWCSISYYELNQRVGETFHA
SQPSMTVDGFTDPSNSERFCLGLLSNVNRNAAVELTRRHIGRGVRLYYIG
GEVFAECLSDNAIFVQSPNCNQRYGWHPATVCKIIPPGCNLKIFNNQEFAA
LLAQSVNQGFEAVYQLTRMCTIRMSFVKGWGAEYRRQTVTSTPCWIELHL
NGPLQWLDKVLTQMGSPSIRCSSVS

Hsmad3
MSSILPFTPPIVKRLLGWKKGEQNGQEEKWCEKAVKSLVKKLKKTGQLDE
LEKAITTQNVNTKCITIPRSLDGRLQVSHRKGLPHVIYCPVRWPDLHSHH
ELRAMELCEFAFNMKKDEVCVNPYHYQRVETPVLPPVLVPRHTEIPAEFP
PLDDYSHSIPENTNFPAGIEPQSNIPETPPPGYLSEDGETSDHQMNHSMD
AGSPNLSPNPMSPAHNNLDLQPVTYCEPAFWCSISYYELNQRVGETFHAS
QPSMTVDGFTDPSNSERFCLGLLSNVNRNAAVELTRRHIGRGVRLYYIGG
EVFAECLSDSAIFVQSPNCNQRYGWHPATVCKIIPPGCNLKIFNNQEFAAL
LAQSVNQGFEAVYQLTRMCTIRMSFVKGWGAEYRRQTVTSTPCWIELHLN
GPLQWLDKVLTQMGSPSIRCSSVS

Figure 12

17/30

```
301/1
ATG TCC TCG ATA CTT CCC TTC ACT CCC CCC ATT GTC AAG AGG CTG CTG GGA TGG AAG AAG GAG CAG AAC GGA
 M   S   S   I   L   P   F   T   P   P   I   V   K   R   L   L   G   W   K   K   E   Q   N   G

331/11                                            361/21
                                                                    CAA GAG GAG AAA TCG
                                                                     Q   E   E   K   W

391/31
TGC GAG AAG GCA GTG AAA AGT CTG GTG AAG AAG AGC GGC CAG CTG GAC GAG CTG GAG AAG GCA CTG
 C   E   K   A   V   K   S   L   V   K   K   K   S   G   Q   L   D   E   L   E   K   A   L
                                                                    451/51    ACC ACC CAG AGC ATC
                                                                            Q   T   T   Q   S   I

481/61
AGC ACC AAG TGC ATC ACC ATC CCC AGG TCT CTG GAT GGG AGA CTC CAG GTG TCC CAT CGT AAG GGG TTG CCT CAT
 S   T   K   C   I   T   I   P   R   S   L   D   D   G   R   L   Q   V   S   H   R   K   G   L   P   H
                              511/71                                      541/81

571/91
CTC TGG AGG TGG CCA GAT CTG CAC AGT CAC CAC GAG CTT CGA GCC ATG GAG GTG TGT GAA TAT GCC TTC AGC ATG
 L   W   R   W   P   D   L   H   S   H   H   E   L   R   A   M   E   V   C   E   Y   A   F   S   M
                              601/101                                     631/111   GTC ATA TAC TGC CGC
                                                                                     V   I   Y   C   R

661/121
TGT GTG AAT CCT TAT CAC TAC CAG AGA GTG GAG ACT CCA GTT CTA CCT CCT GTG CTG CCT GTG GTT AAG AAG GAC GAA GTG
 C   V   N   P   Y   H   Y   Q   R   V   E   T   P   V   L   P   P   V   L   P   P   V   L   V   K   K   D   E   V
                  691/131                                      721/141   CCA CGA AAC ACC GAA
                                                                          P   R   N   T   E

751/151
CCT TCA CTT GAT GAT TAC AGC CAT CCC ATT CCG GAA AAT ACC AAT TTC CCA GCA GGC ATC GAG CCA GAA
 P   S   L   D   D   Y   S   H   I   P   E   N   T   N   F   P   A   G   I   E   P   A   E   F
                              781/161                                     811/171   ATT CCG GCA GAA TTC
                                                                                     I   P   A   E   F

841/181
CCT CCT CCC GGT TAC TTG AGC GAA GAT GGG GAA ACA AGC GAC CAA ATG AAT CAC AGT ATA GAC ACG TAC ATT CCA GAA ACG
 P   P   P   G   Y   L   S   E   D   G   E   T   S   D   Q   M   N   H   S   I   D   T   Y   I   P   E   T
                  871/191                                      901/201   TAC ATT CCA GAA ACG
                                                                          Y   I   P   E   T

931/211
TCT ATG TCT CCT GCT CAT AGC AAC ATG GAC CTG CAG CCT GTC ACA TAC TGC GAG CCG GCC TTT TGG AAT CTG TCA CCA AAC
 S   M   S   P   A   H   S   N   M   D   L   Q   P   V   T   Y   C   E   P   A   F   W   C   S   I   N   L   S   P
                  961/221                                      991/231   AAT CTG TCA CCA AAC
                                                                          N   L   S   P

1021/241
AAC CAA CGC GTA GGG GAG ACG TTC CAC GCT TCC CAG CCC TCC ATG ACA GTG GAT GGA TTC ACC GAT CCT TCC AAC
 N   Q   R   V   G   E   T   F   H   A   S   Q   P   S   M   T   V   D   G   F   T   D   P   S   N
                  1051/251                                     1081/261   TCC TAC TAT GAG CTT
                                                                           S   Y   Y   E   L
                                                                TCT GAA CGT TTC TGC
                                                                 S   E   R   F   C
```

Figure 12

```
1111/271
CTG GGG CTG TTG TCC AAC GTA AAT CGG AAT GCA GCT GTG GAG CTG ACA CGG AGA CAC ATC GGG AGA GGC GTG CGG CTG
 L   G   L   L   S   N   V   N   R   N   A   A   V   E   L   T   R   R   H   I   G   R   G   V   R   L
                                 1141/281                                      1171/291

1201/301                                                                                       1261/321            TAT TAC ATT GGA
GGG GAA GTG TTT GCC GAG TGC CTC AGT GAC AAT GCC ATA TTT GTA CAG TCC CCA AAT TGT AAC CAG TAC GGT TGG          Y   Y   I   G
 G   E   V   F   A   E   C   L   S   D   N   A   I   F   V   Q   S   P   N   C   N   Q   Y   G   W
                                 1231/311

1291/331                                                                                       1351/351            CAT CCT GCC ACA
GTC TGC AAG ATT CCA CCA GGC TGT AAC CTG AAG ATA TTT AAT AAC CAG GAG TTT GCT GCT CTT TTG GCT CAG TCA GTA          H   P   A   T
 V   C   K   I   P   P   G   C   N   L   K   I   F   N   N   Q   E   F   A   A   L   L   A   Q   S   V
                                 1321/341

1381/361                                                                                       1441/381            AAC CAG GGC TTT
GAG GCT GTG TAT CAG CTT ACG AGG ATG TGC ACC ATA CGC ATG AGT TTC GTC AAA GGC TGG GGA GCC GAA TAC AGG CGA          N   Q   G   F
 E   A   V   Y   Q   L   T   R   M   C   T   I   R   M   S   F   V   K   G   W   G   A   E   Y   R   R
                                 1411/371

1471/391                                                                                       1531/411            CAG ACT GTG ACT
AGC ACC CCC TGC TGG ATC GAG CTG CAC CTG AAC GGG CCC TTG CAA TGG TTG GAT AAG GTT CTC ACT CAG ATG GGG TCT          Q   T   V   T
 S   T   P   C   W   I   E   L   H   L   N   G   P   L   Q   W   L   D   K   V   L   T   Q   M   G   S
                                 1501/401

1561/421                                                                                                                             CCA AGT ATC CGC
TGC TCC AGT GTT TCT TAA                                                                                                                P   S   I   R
 C   S   S   V   S   .
```

Figure 12

A

Mixer	KH-SQMPFHPSLLMDENNFPPNKTITPDMNV--RIPPIPVSAPSNNHS
ZF Mixer	ANHAKSTMKQFLV-EYDNFPPNKTIGPEMKV--VIPPLPSQSNFMMSS
Milk	RH-NQVSMHSNLMMDF---PPNKTITPDMNT--IIPQITDATGWSSQE
Bix3	RQ-NQVTMHSNLVMEF---PPNKTITPDMNT--IIPQIPGATGWKNQE
XFast1	DSPRGPQSPLDLDNMLRAMPPNKSV-FDVLT--SHPGDLVHPSFLSQC
HFast1	ETRGPPGLLCDLDALFQGVPPNKSI-YDVWV--SHPRDLAAPGPGWLL
MFast2	ESQGSQDLLCDLDSLFQGVPPNKSI-YDVWV--SHPRDLAAPAPGWLL
XFast3	PQIPLTPKPPELKNAPSDFPPNKTV-FDIPVYTGHPGFLASQSLFSPH

Figure 13

B

Mixer SIM peptide

Mixer	SQMPFHPSLLMDFNNFPPNKTITPDMNVRIPPIPVSAPSNNH-
ZF Mixer	AKSTMKQFL-VEYDNFPPNKTIGPEMKVVIPPLPSQSNFMMSS
Milk	NQVSMHSNLMMDF---PPNKTITPDMNTIIPQITDATGWS---
Bix3	NQVTMHSNLVMEF---PPNKTITPDMNTIIPQIPGATGWK---
Bix1	NQVSMHSNLMMDISNFPPTKTITANMNTIIPQMPGASCWS---
Bix4	NQVSMHSNIMMDFSNFQPKKTVTPDMNTIIPQIPDATGWS---
Mix.1	QQMPVQPMLMNSF---QTNKNIKPEVYTTSPQIPVSTTS----
Mix.2	QQMPAQPMFMNSF---QTNKIIKSKMDTTSPPIPVSTTSSHH-

SRMNVFNTKEAGPLVSLPEDVYEE	371
SSPKHIACSVQNMSVQTQPELFAT	327
SQEGTDAYSTQGALPRAQCSPYGQ	400
NQEDINTYSTQGALSRAGCSSYGL	382
SHEDINAYSTQGAVPMAGCSPYGH	401
NQEGTDAYSTQEALPRAQCSPYGH	380
SQVSLFANQEPCHMSTTQGGTYGQ	377
SQMSLFAGQDPCHMSTAPGGTYGQ	369

Figure 13

C   XFast3

MSFGLHPWDVAFRPSPPHNLEKKVVPPGADREKSLPSPKEDSDGAREPDSTVDLRKKNKKKNYQ

RYAKPPYSYLAMISLVIQNSPEKRLKLSQILQDISSLFPFFKGNYQGWKDSIRHNLSSNDCFRKV

LKDPLKPQAKGNYWTVDVTRIPPDALKLQNTAVTRQDLFPLDLAPYILHGQPYRSLERLSANHTR

GRTTPRMEPEVQIPVSDPAVSFPMILWNLPTSYSKCVAPNVVAPPSIHPLLLYSNFPSISIYNYL

PPPYGSPVYSDRRDLLASGLHPQIPLTPKPPPELKNAPSDFPPNKTVFDIPVYTGHPGFLASQSLF

SPHLPTATPPLVGYRPSGL

Figure 13

Figure 14

Figure 15
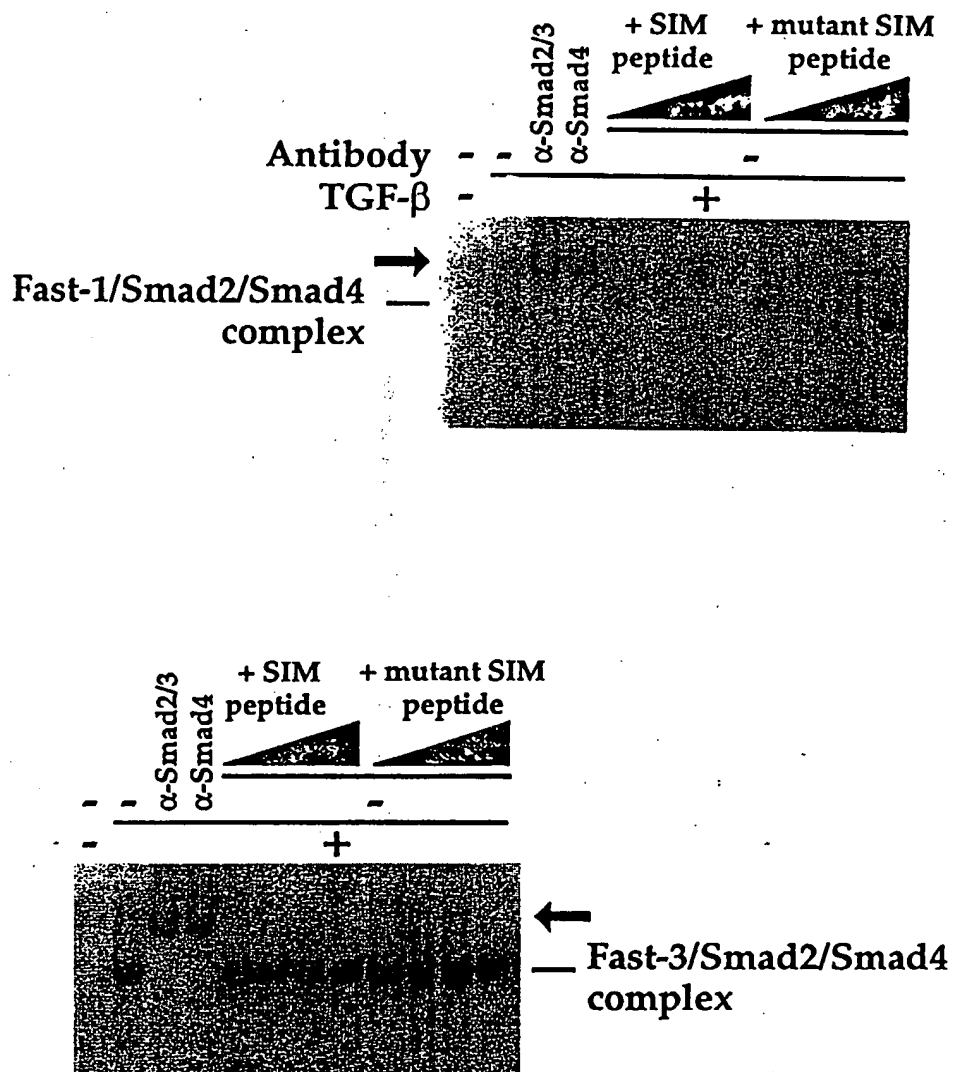
Figure 15

Figure 16

B

Figure 16

27/30

Figure 17

C

+ mutant SIM peptide +
+ SIM peptide +
+
-

TGF-β —

JunB —

γ-actin —

Figure 17

B

+ mutant SIM peptide
+ SIM peptide
α-Smad2/3 + α-Smad4
α-Smad4
α-Smad2/3
Antibody
TGF-β
XFast-3

Fast-3/Smad2/Smad4
complex

Free Probe —
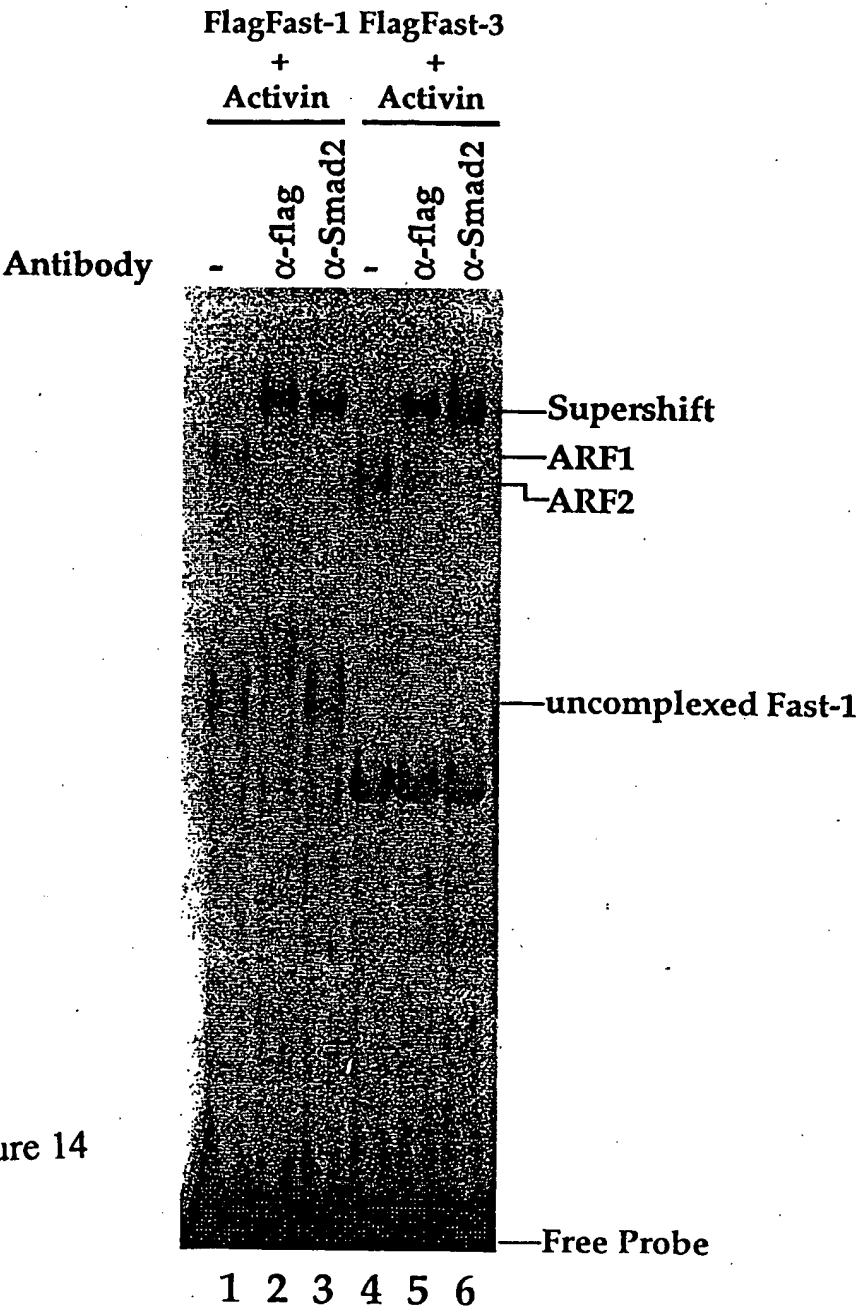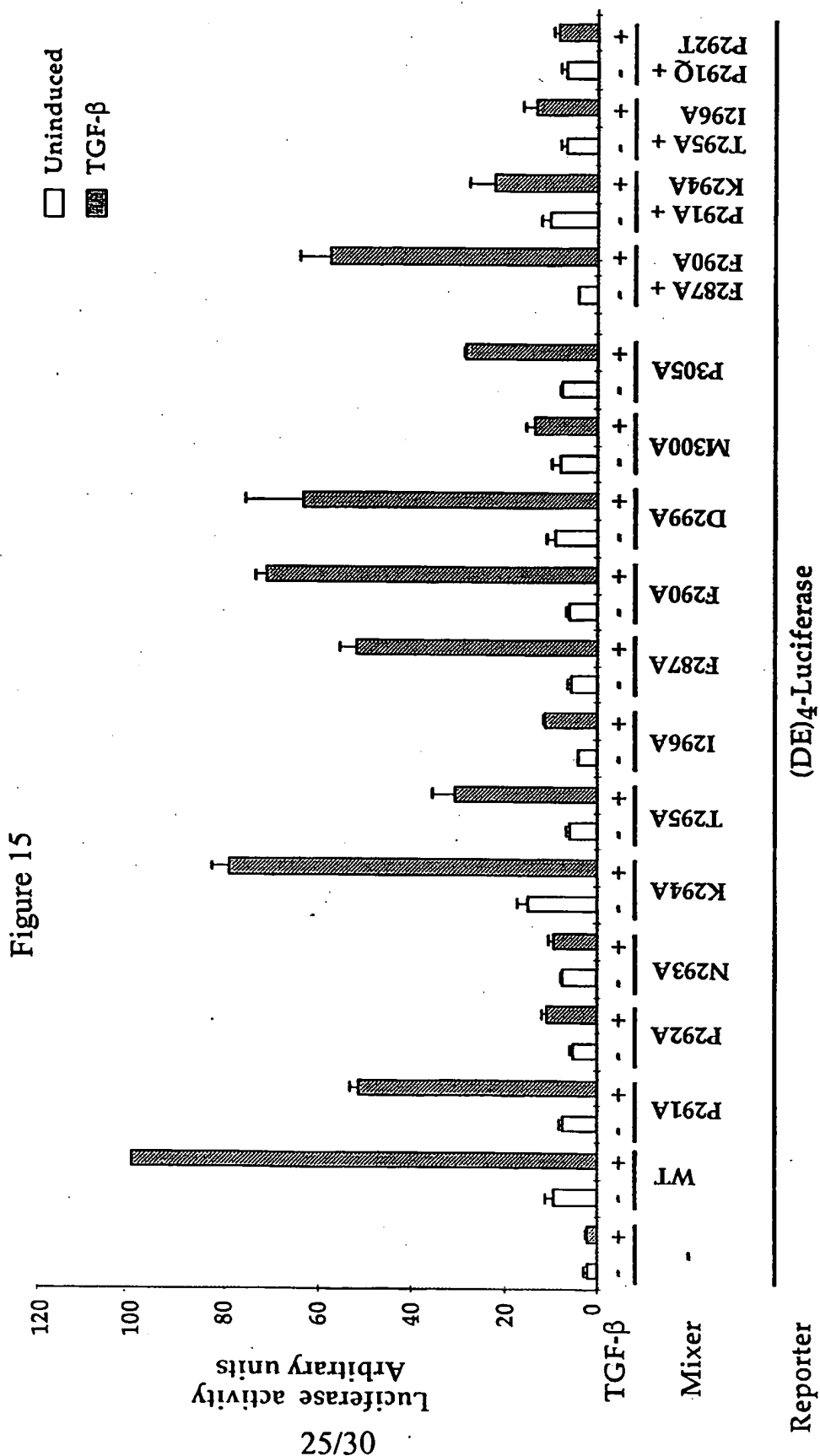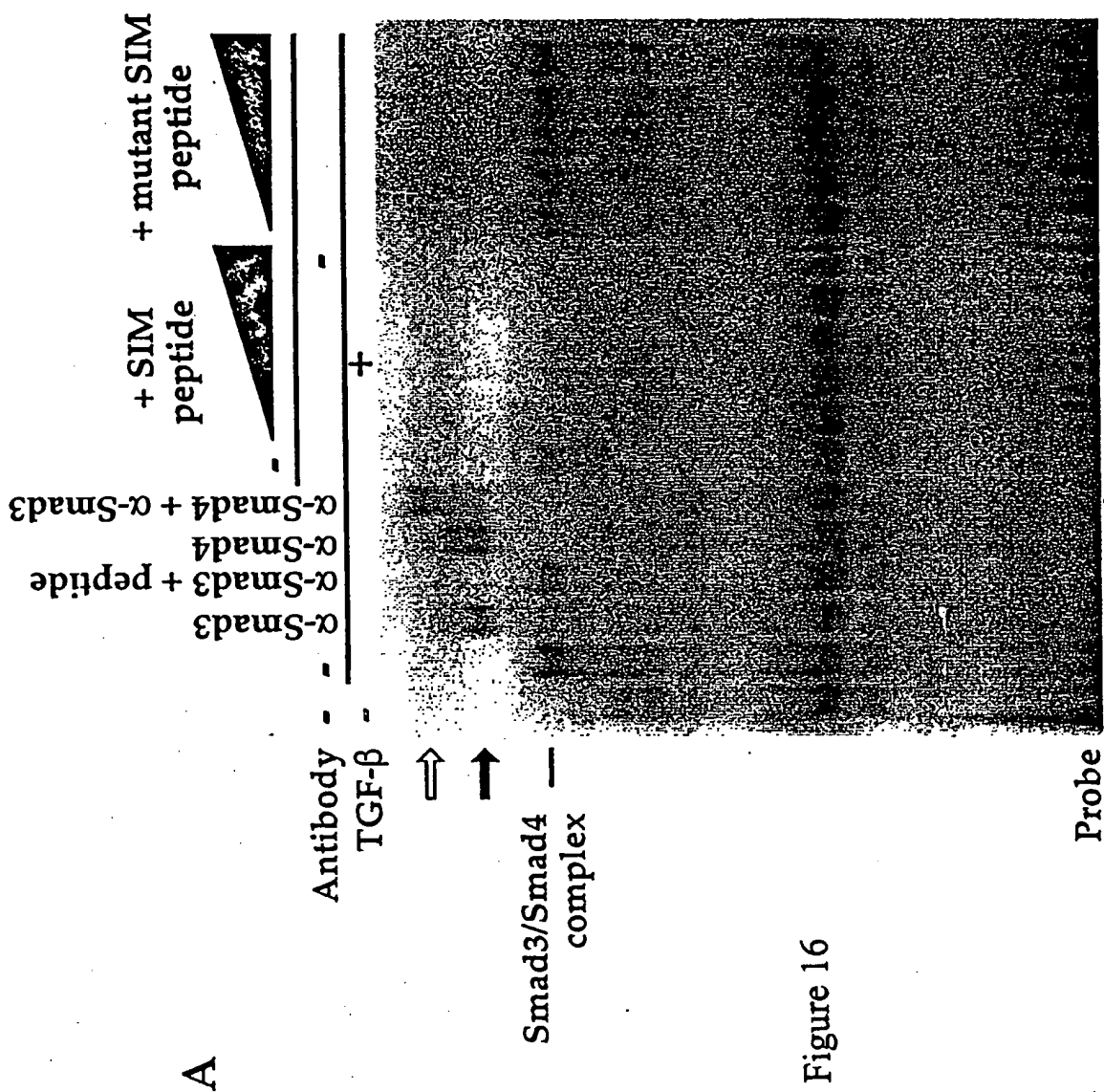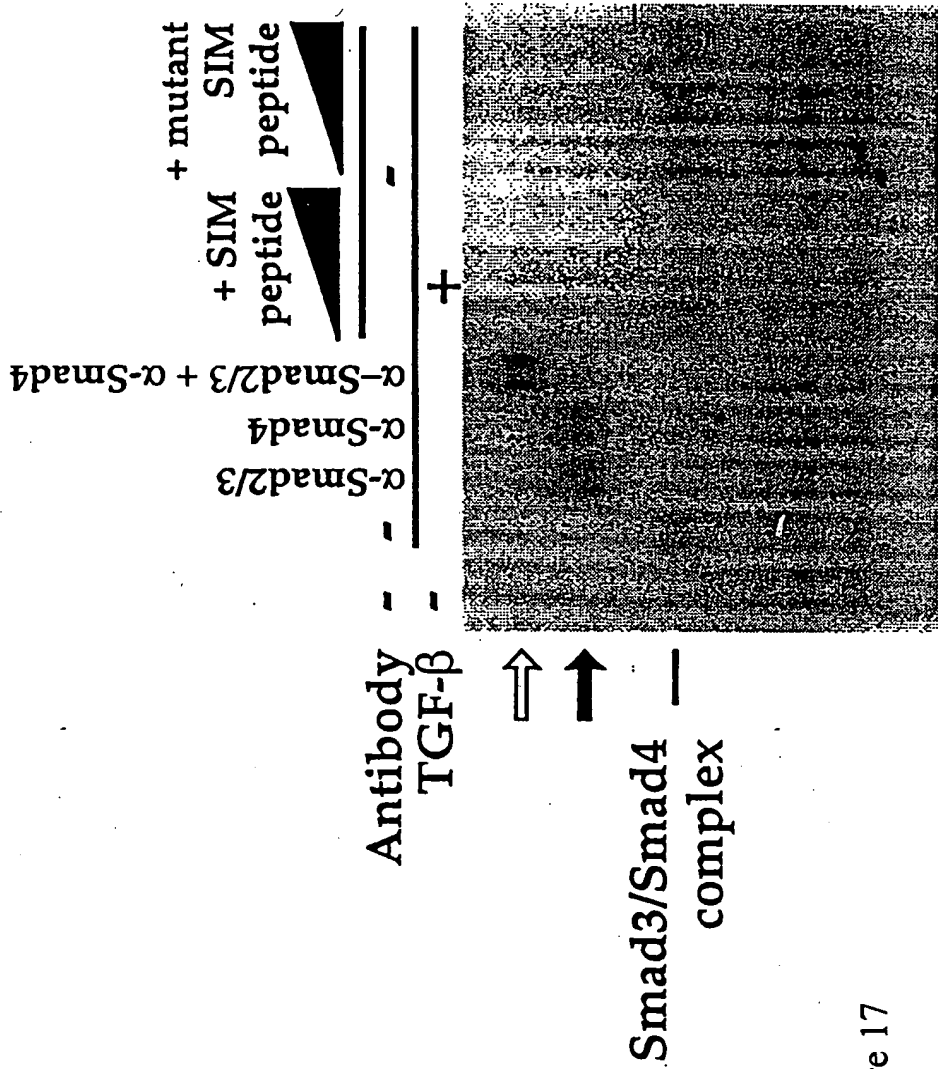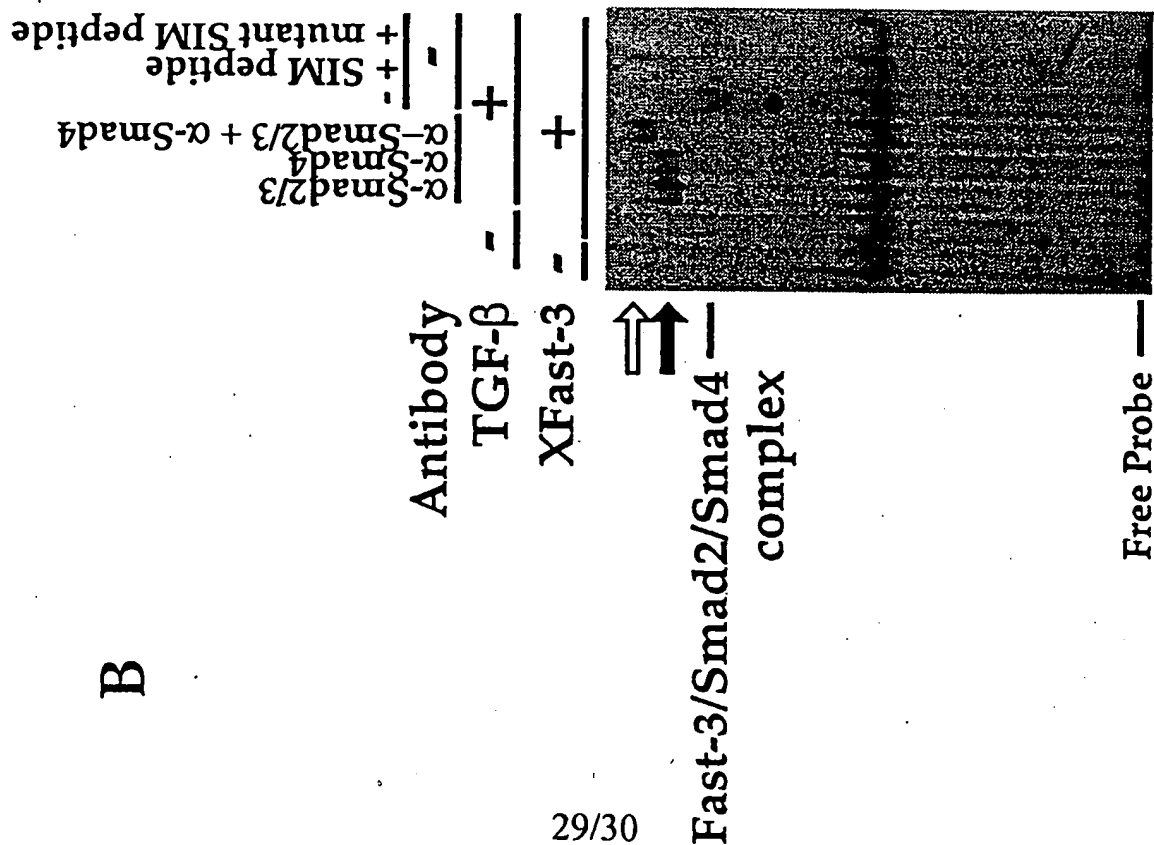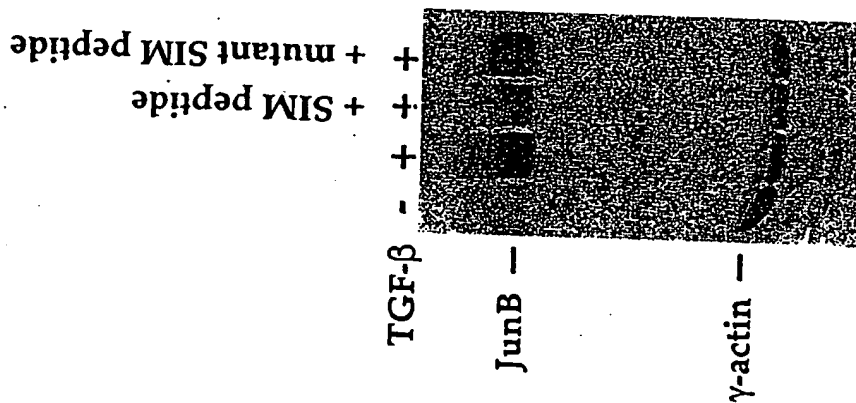
Figure 18

```
   1  ATGTCTTTG  GGCTTCACCC  ATGGGATGTG  GCCTTCAGAC  CTTCACCCCC
      TCACAACCTG  GAGAAGAAAG  TCGTCCCCCC  AGGGGCCGAC  AGAGAGAAGT

 101  CGCTTCCTTC  CCCAAGGAG  GACAGTGATG  GGGCCCGGGA  GCCCGACTCC
      ACTGTGGATT  TGAGGAAGAA  GAACAAGAAG  AAGAAGAACT  ACCAGAGATA

 201  CGCCAAGCCC  CCCTATTCCT  ACCTGGCCAT  GATCTCCCTG  GTCATCCAGA
      ACTCCCCCGA  GAAGAGGCTC  AAACTCTCCC  AGATCCTGCA  GGACATCAGC

 301  TCTCTGTTTC  CATTCTTCAA  GGGCAACTAC  CAGGGCTGGA  AGGATTCCAT
      TCGGCATAAT  TTGTCTTCCA  ACGACTGTTT  CAGAAAGGTT  CTGAAGGATC

 401  CGCTCAAGCC  ACAGGCCAAG  GGCAATTACT  GGACAGTAGA  CGTGACCCGG
      ATCCCCCCAG  ACGCTTTGAA  GCTCCCAGAAC  ACGGCGGTGA  CCCGGCAGGA

 501  CCTGTTCCCC  CTGGACCCTGG  CCCCCTACAT  CCTACACGGG  CAGCCGTACA
      GGAGTCTGGA  GAGGCTCTCG  GCCAATCACA  CGAGGGGGCG  CACGACCCCC

 601  AGGATGGGAGC  CTGAAGTTCA  GATTCCAGTG  TCAGACCCAG  CTGTCAGTTT
      CCCCATGATC  CTATGGAATC  TGCCGACATC  CTACAGCAAA  TGTGTGGCCC

 701  CCAATGTAGT  GGCCCCTCCC  AGCATTCACC  CCCTCTTGTT  GTACTCCAAC
      TTCCCTTCCA  TTTCCATTTA  TAACTACCTG  CCCCCGCCCT  ATGGCAGCCC

 801  CGTGTACTCA  GACAGACGAG  ATCTTCTTGC  CTCCGGCCTG  CACCCCCAAA
      TCCCTCTCAC  CCCCAAACCC  CCAGAGCTGA  AGAACGCCCC  CAGCGACTTC

 901  CCCCCCAACA  AGACAGTGTT  TGACATCCCC  GTCTATACTG  GCCACCCGGG
      GTTCCTTGCT  AGCCAAAGCT  TGTTCAGCCC  ACACTTGCCC  ACGGCTACAC

1001  CCCCCCTCGT  GGGCTACCGG  CCATCTGGGC  TATGA
```

30/30